DOCUMENT RESUME

ED 442 840 TM 031 260

AUTHOR	McMillan, James H.; Workman, Daryl
TITLE	Teachers' Classroom Assessment and Grading Practices: Phase I and II.
INSTITUTION	Metropolitan Educational Research Consortium, Richmond, VA.
PUB DATE	1999-00-00
NOTE	145p.
PUB TYPE	Reports - Evaluative (142) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; Educational Practices; Elementary Secondary Education; Grades (Scholastic); *Grading; Performance Factors; Questionnaires; *Student Evaluation; Teacher Surveys; *Teachers
IDENTIFIERS	Virginia

ABSTRACT

Teacher assessment and grading practices were studied in a two-phase investigation in the seven school districts that make up the Metropolitan Educational Research Consortium in Virginia. The first report summarizes the findings from Phase 1 of the study, which focused on teacher responses to closed-end, written survey questions. In Phase 1, 921 elementary, 597 middle, and 850 high school teachers were surveyed. They were asked about their grading and classroom assessment practices for a "typical" first semester class. Elementary school teachers indicated that academic factors clearly are most important in determining grades, but that related factors, such as improvement, effort, ability level, and class participation also make a significant contribution. The variety of responses shows large differences in how teachers emphasize different factors. Approximately 20% of grades given were "A"s. Results for secondary school teachers show little variation between grade levels or subject matter. As with elementary school teachers, academic performance was the most important grading factor, but effort, homework, and extra credit also entered into grading. Phase 2 of the study focuses on interviews with 28 teachers. The analysis of interview data indicates that there is tension between two sources of influence on teacher decision-making concerning assessment and grading practices. One source is teacher beliefs and values and another is external pressures and constraints. These pressures include parent demands and informing parents of student progress, school division policies, skills needed by students once they graduate, practical constraints and state-mandated high-stakes multiple-choice testing. The state test seems to have become a significant influence on teacher decision making. An appendix to Phase I contains the teacher surveys. (Contains 20 tables and 60 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 840

TEACHERS' CLASSROOM ASSESSMENT AND GRADING PRACTICES:

Phase I and II

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

TM031260

G.D. Hipp

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CHESTERFIELD COUNTY PUBLIC SCHOOLS • COLONIAL HEIGHTS CITY SCHOOLS • HANOVER COUNTY PUBLIC SCHOOLS • HENRICO COUNTY PUBLIC SCHOOLS • HOPEWELL CITY PUBLIC SCHOOLS • POWHATAN COUNTY PUBLIC SCHOOLS • RICHMOND CITY PUBLIC SCHOOLS • VIRGINIA COMMONWEALTH UNIVERSITY

BEST COPY AVAILABLE

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

PO BOX 842020 • RICHMOND VA 23284-2020 • Phone: (804) 828-0478
FAX: (804) 828-0479 • E-MAIL: jmcmilla@saturn.vcu.edu

MERC MEMBERSHIP
James H. McMillan, Director

CHESTERFIELD COUNTY PUBLIC SCHOOLS
William C. Bosher, Jr., Superintendent

COLONIAL HEIGHTS CITY SCHOOLS
James L. Ruffa, Superintendent

HANOVER COUNTY PUBLIC SCHOOLS
Steward D. Roberson, Superintendent

HENRICO COUNTY PUBLIC SCHOOLS
Mark A. Edwards, Superintendent

HOPEWELL CITY PUBLIC SCHOOLS
David C. Stuckwisch, Superintendent

POWHATAN COUNTY PUBLIC SCHOOLS
Margaret S. Meara, Superintendent

RICHMOND CITY PUBLIC SCHOOLS
Albert J. Williams, Superintendent

VIRGINIA COMMONWEALTH UNIVERSITY
John S. Oehler, Dean
School of Education

Virginia Commonwealth University and the school divisions of Chesterfield, Colonial Heights, Hanover, Henrico, Hopewell and Richmond established the **Metropolitan Educational Research Consortium** (MERC) on August 29, 1991. The founding members created **MERC** to provide timely information to help resolve educational problems identified by practicing professional educators. **MERC** membership is open to all metropolitan-type school divisions. It currently provides services to 9,000 teachers and 138,000 students. **MERC** has base funding from its membership. Its study teams are composed of University investigators and practitioners from the membership.

MERC is organized to serve the interests of its members by providing tangible material support to enhance the practice of educational leadership and the improvement of teaching and learning in metropolitan educational settings. **MERC's** research and development agenda is built around four goals:

- To improve educational decision-making through joint development of practice-driven research questions, design and dissemination,
- To anticipate important educational issues and provide leadership in school improvement,
- To identify proven strategies for resolving instruction, management, policy and planning issues facing public education, and
- To enhance the dissemination of effective school practices.

In addition to conducting research as described above, **MERC** will conduct technical and issue seminars and publish reports and briefs on a variety of educational issues.

TEACHERS' CLASSROOM ASSESSMENT AND GRADING PRACTICES:

Phase I

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

CHESTERFIELD COUNTY PUBLIC SCHOOLS • COLONIAL HEIGHTS CITY SCHOOLS • HANOVER COUNTY PUBLIC SCHOOLS • HENRICO COUNTY PUBLIC SCHOOLS • HOPEWELL CITY PUBLIC SCHOOLS • POWHATAN COUNTY PUBLIC SCHOOLS • RICHMOND CITY PUBLIC SCHOOLS • VIRGINIA COMMONWEALTH UNIVERSITY

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

PO BOX 842020 • RICHMOND VA 23284-2020 • Phone: (804) 828-0478
FAX: (804) 828-0479 • E-MAIL: jmcmilla@saturn.vcu.edu

MERC MEMBERSHIP
James H. McMillan, Director

CHESTERFIELD COUNTY PUBLIC SCHOOLS
William C. Bosher, Jr., Superintendent

COLONIAL HEIGHTS CITY SCHOOLS
James L. Ruffa, Superintendent

HANOVER COUNTY PUBLIC SCHOOLS
Steward D. Roberson, Superintendent
Chairman, MERC Policy & Planning Council

HENRICO COUNTY PUBLIC SCHOOLS
Mark A. Edwards, Superintendent

HOPEWELL CITY PUBLIC SCHOOLS
David C. Stuckwisch, Superintendent

POWHATAN COUNTY PUBLIC SCHOOLS
Margaret S. Meara, Superintendent

RICHMOND CITY PUBLIC SCHOOLS
Albert J. Williams, Superintendent

VIRGINIA COMMONWEALTH UNIVERSITY
John S. Oehler, Dean
School of Education

Virginia Commonwealth University and the school divisions of Chesterfield, Colonial Heights, Hanover, Henrico, Hopewell and Richmond established the **Metropolitan Educational Research Consortium** (MERC) on August 29, 1991. The founding members created **MERC** to provide timely information to help resolve educational problems identified by practicing professional educators. **MERC** membership is open to all metropolitan-type school divisions. It currently provides services to 9,000 teachers and 138,000 students. **MERC** has base funding from its membership. Its study teams are composed of University investigators and practitioners from the membership.

MERC is organized to serve the interests of its members by providing tangible material support to enhance the practice of educational leadership and the improvement of teaching and learning in metropolitan educational settings. **MERC's** research and development agenda is built around four goals:

- To improve educational decision-making through joint development of practice-driven research questions, design and dissemination,
- To anticipate important educational issues and provide leadership in school improvement,
- To identify proven strategies for resolving instruction, management, policy and planning issues facing public education, and
- To enhance the dissemination of effective school practices.

In addition to conducting research as described above, **MERC** will conduct technical and issue seminars and publish reports and briefs on a variety of educational issues.

TEACHERS' CLASSROOM ASSESSMENT AND GRADING PRACTICES:

Phase I

James H. McMillan, Professor
Virginia Commonwealth University

Daryl Workman, MERC Research Fellow
Virginia Commonwealth University

November 1998

Copyright© 1998. Metropolitan Educational Research Consortium (MERC), Virginia Commonwealth University

*The views expressed in MERC publications are those of individual authors and not necessarily those of the Consortium or its members.

Executive Summary

Classroom assessment and grading practices are becoming a greater focus of educational inquiry as teachers and policymakers become more accountable to the public for educational outcomes via assessment driven instructional practices. This study was an attempt to better understand the classroom assessment and grading practices of teachers, which have previously been described as a "hodgepodge" mix of student attitude, effort and achievement. Specifically, the following questions regarding teachers' assessment and grading practices were addressed:

- What is the current state of assessment practice and grading by teachers?
- What assessment and grading topics do teachers identify as needs to be addressed in in-service?
- What is the relationship between assessment and grading practices and grades given to students?
- What are the relationships between grade level, and subject taught and assessment and grading practices?
- What are the reasons teachers give for their assessment and grading decision-making?
- What is the impact of the SOL tests on the extent to which different assessment techniques are used in the classroom?
- What classroom assessment and in-service needs do teachers have?

Results of the study indicate that teachers do in fact use a multitude of factors to assess and grade students, including academic performance, effort, improvement, ability, homework, and extra credit. However, this study looked beyond a "hodgepodge" explanation of assessment and grading practices to uncover relationships that help to further explain teachers' assessment and grading practices and decision-making processes.

This report summarizes the findings from Phase I of the study, which focused on teacher responses to closed-end, written survey questions. Phase II summarizes personal interviews with teachers. In Phase I, 921 elementary, 597 middle and 850 high school teachers were surveyed. The teachers were asked questions about their grading and classroom assessment practices for a "typical" first semester class.

Elementary teachers indicated that academic factors clearly are most important in determining grades, but that related factors, such a improvement, effort, ability level and class participation also make a significant contribution. However, there is a high

variation of responses among the teachers, showing large differences in how much different teachers emphasize different factors. Approximately 20% of grades given are As. Elementary teachers indicated that many classroom assessment topics need staff development. Following data reduction analyses, relationships between content areas (math and language arts) and grade level were examined. Few relationships were found.

Results from secondary teachers showed little variation between grade levels or subject matter.

Like elementary teachers, there were four major factors used in grading: academic performance, academic-enablers such as effort, homework, and extra credit. Clearly academic performance is most influential, but academic enabling behaviors are also very important, especially for some teachers. Also like elementary teachers, there is great variation among the teachers in the weight given to different factors, suggesting an idiosyncratic approach to grading. There is great reliance on teacher-made tests. Essay and objectives tests are used about the same, and there is extensive use of constructed-response assessments such as performance assessments and projects. Advanced classes emphasize academic performance, constructed response items, major exams, and reasoning more than standard of basic classes. Few other relationships were found between grade level or subject taught and assessment or grading practices.

Implications of the findings are discussed, including the need for clarifying how academic enabling factors are incorporated, whether idiosyncratic practices should be maintained, effects of SOL testing, needs for professional development, use of zeros in calculating grades, and how differentiation of so-called "higher order" thinking skills are differentiated from recall and understanding. Further analyses of the findings will be done when Phase 2 of the research is completed.

Preface

The research in this report was directed by a team of individuals. This team identified the research problem and questions, developed a research design, assisted in gathering data from teachers, and took an active role in identifying samples and analyzing data. The principal investigators are grateful for their contribution and assistance. The members of the research team include the following:

James McMillan
Virginia Commonwealth University

Daryl Workman
Virginia Commonwealth University

Lin Corbin-Howerton
Chesterfield County Public Schools

Joseph Tylus
Monacan High School
Chesterfield County Public Schools

Ann Williams
Colonial Heights Middle School
Colonial Heights Public Schools

James Bagby
Hanover County Public Schools

Carole Urbansok-Eads
Hanover County Public Schools

Deborah Pittman
Cold Harbour Elementary School
Hanover County Public Schools

Catherine Nolte
Henrico County Public Schools

Yvonne Smith-Jones
Hopewell City Public School

Stephanie Couch
Pocahontas Middle School
Powhatan County Public Schools

Charmaine Brooks
Westover Hills Elementary
Richmond City Public Schools

Pat Janes
Armstrong High School
Richmond City Public Schools

Sue Jones
Patrick Copeland Elementary
Richmond City Public Schools

Audrey Johnson
Thomas Jefferson High School
Richmond City Public Schools

Richard Williams
Richmond City Public Schools

Table of Contents

Introduction 1

Review of Literature 3

Research Questions 12

Methodology 13
- Research Design 13
- Population and Sample 14
- Instrument 14
- Procedure 16
- Data Analyses 17

Elementary Findings 17
- Descriptive Results 17
- Grades 22
- In-Service Needs 23
- Data Reduction 29
- Relationship Results 33
- Summary of Elementary Level Findings 36

Secondary Findings 37
- Descriptive Results 39
- In-Service-Needs 44
- Data Reduction 49
- Relationship Results 51
- Summary of Secondary Level Findings 58

Conclusions 60

Implications 62

References 64

Appendices 67

Introduction

A significant amount of recent literature has focused on classroom assessment and grading as essential aspects of effective teaching. There is an increased scrutiny of assessment as indicated by the popularity of performance assessment and portfolios, newly established national assessment competencies for teachers (Standards, 1990), and the interplay between learning, motivation, and assessment (Brookhart, 1993, 1994; Tittle, 1994). In Virginia, the Standards of Learning and associated tests highlight the importance of assessment.

Previous research documents that teachers tend to award a "hodgepodge grade of attitude, effort, and achievement" (Brookhart, 1991, p. 36). It is also clear that teachers use a variety of assessment techniques, even if established measurement principles are often violated (Cross & Frary, 1996; Frary, Cross, & Weber, 1993; Plake & Impara, 1993; and Stiggins & Conklin, 1992).

Given the variety of assessment and grading practices in the field, the increasing importance of assessment, the critical role each classroom teacher plays in determining assessments and grades, and the trend toward greater accountability of teachers with state assessment approaches that are inconsistent with much of the current literature, there is a need to (1) understand current assessment and grading practices, (2) understand the relationship of these practices to grades given by teachers, (3) determine if "standards" teachers use to assign grades differ from one classroom to another and one school to another, (4) examine the consequential validity of the new SOL tests on classroom assessment practices, and (5) determine assessment and grading topics that, according to teachers, need in-service.

The fourth need is related to a recently expanded conception of test validity that includes what has been called "consequential validity" or "consequential bias" (Messick, 1989; Moss, 1992). Essentially, test developers and users need to be sensitive to how assessments influence instructional practices and curriculum. The importance of consequential validity is indicated by its inclusion in the new Standards for Educational and Psychological Testing. Of interest in the current study is the effect the new statewide assessment program may have on instructional practices. For example, the assessments may result in teachers stressing a particular method of instruction or classroom testing that is consistent with the emphasis and approach adopted in the statewide system

There is a need to provide information that addresses issues of consistency and fairness in assessment and grading across classrooms and schools, to illustrate to teachers the nature of current practice and provide a stimulus for discussion, and to establish assessment and grading policy. There is also a need to understand the motivation and reasons for using specific assessment and grading practices.

The purpose of this investigation, then, was to describe the classroom assessment and grading practices of teachers, determine if meaningful relationships exist between these practices and grade level, subject matter, ability levels of different classes, and to understand the reasons teachers give for using certain assessment and grading practices, and to document teacher needs for inservice education related to assessment.

Review of Literature

Despite the growing importance of classroom assessment and the introduction of new methods of assessment, there is relatively little research on the nature and effects of classroom assessments on student learning and motivation (Stiggins, 1997). Most assessment research has focused on standardized testing, despite evidence that teachers spend considerable time assessing students, and that student well-being is influenced by the quality of assessments given by the teacher (Stiggins and Conklin, 1992). Also, there is little empirical research on classroom assessments, with measurement experts tending instead to pay much more attention to large scale testing than classroom assessment. It is also evident that many teachers lack assessment competency (Plake and Impara, 1997). This isn't too surprising, however, since less than 50% of the teacher certification programs in the United States require no measurement course (Schafer, 1993). This remains the case, despite the fact that teacher standards for assessment competency were identified in 1990 (AFT, NCME, NEA, 1990).

Prior to the mid 1980s the literature on educational assessment focused almost exclusively on large-scale standardized testing. According to Stiggins and Conklin (1992), most inquiry on classroom assessment was based on a conceptualization similar to what had been developed for standardized testing, emphasizing paper and pencil, multiple choice testing. Furthermore, the only written standards for assessment, *Standards for Educational and Psychological Testing*, dealt primarily with standardized tests. Finally, during the 1980s the emerging literature about teacher decision-making, teacher behavior, and student achievement

found little on how classroom assessments relate to teaching or learning. Shulman (1980) concluded that most of the paper and pencil tests used for assessment were inconsistent with, and often irrelevant to, the realities of teaching. Haertel, et al. (1984), in a review of research on high school testing, concluded that little is known about teachers' or students' perceptions of the impacts of classroom assessment.

Phye (1997) states that "it is not only the assessment option that determines what we get as evidence of learning or achievement. How we use the assessment instruments or techniques also determine the nature of the knowledge a student is demonstrating. *How* we assess determines *what* we get" and thus classroom learning and assessment "go hand in hand" (p.51).

Airasian (1984) reviews literature that suggests teachers focus their classroom assessments in two areas: academic achievement and social behavior. The importance of these factors varies with grade level, with elementary teachers placing greater importance on social behavior. Airasian also found that teachers' informal "sizing up" assessments remain relatively stable throughout the year and influence student self-perceptions of ability.

Fleming and Chambers (1983), in a study that analyzed nearly 400 teacher-developed classroom tests, came to several conclusions:

- Short-answer questions are used most frequently.
- Essay questions are avoided, representing slightly more than 1% of test items.
- Matching items are used more than multiple choice or true false items.
- Most test questions, approximately 80%, sample knowledge of terms, facts, and rules and principles (94% for middle school teachers, 69% for high school teachers, and 69% of elementary school teachers).
- Few test items measure student ability to apply what they have learned.

Research by Carter (1984), in which the test development skills of high school teachers were studied, in support of what Fleming and Chambers found, reported that the teachers had

considerable difficulty recognizing or writing items that tapped "higher order" thinking skills, such as application. Stiggins and Conklin (1992), with a sample of thirty-six teachers, found that recall knowledge items were used approximately fifty percent of the time. There is ample evidence to suggest that many teachers do not have sufficient knowledge and skill to develop, apply, and summarize classroom assessments. In a survey of 228 teachers from four grades (2, 5, 8, and 11), Stiggins and Conklin (1992) report that nearly three fourths of the teachers indicated some concern about their own tests. Examples of the kinds of concerns expressed included: "Are my tests effective? How can I make them better? Do they focus on students' real skills? Are they challenging enough? Do they aid in learning?" (p. 39). Concern was greatest for high school teachers. Only 15% of high school teachers indicated that they had no concerns about their assessments. Stiggins and Conklin also asked 24 teachers to keep a journal to reflect upon their assessment practices. The analysis focused on how teachers describe their assessments and what specific issues were raised related to their assessments. They found that teachers were most interested in assessing student mastery or achievement, and that performance assessment was used frequently. The nature of the assessments used in each class was coupled closely with the roles each teacher set for her students, teacher expectations, and the type of teacher-student interactions desired. The results of these investigations led to the development of classroom assessment profiles. The profile was tested with eight high school classrooms, resulting in the following key factors:

- Assessment purposes
- Assessment methods
- Criteria used in selecting assessment methods
- Quality of assessments
- Feedback to students
- Teacher as assessor (background, preparation)

- Teacher perception of the students
- The assessment-policy environment

These components can be used to characterize diverse assessment practices and environments. Two recent studies document teacher beliefs and knowledge about classroom assessment. Frary, Cross, and Weber (1993) used a statewide random sample of 536 high school teachers of academic subjects to survey self-report practices and beliefs about classroom assessment. Frequency of use of various kinds of test questions revealed the following percentages:

Type of Question	Seldom or never	Frequently or always
Short answer	17%	56%
Essay	41%	38%
Multiple choice	21%	52%
True-false	47%	19%
Performance	30%	37%

These results suggest that teachers use a variety of assessment approaches. The teachers were asked to indicate degree of agreement to many statements concerning grading and assessment practices. Concerning assessment, it was noteworthy that 66% of the teachers agreed that essay tests provide a better assessment of student knowledge than do multiple choice tests; that 47% agreed that the nature of multiple choice items encourages superficial learning; and that better measurement occurs when teachers award partial credit rather than scoring simply right or wrong.

A second survey of teachers, taken in 1992, was structured to obtain teacher competency concerning assessment practices by asking teachers to indicate which of several possible answers to assessment questions was best (Plake and Impara, 1997). A national random sample of 555

elementary, middle, and high school teachers was used. Overall mean performance on the survey was 66% correct. Teachers did better on items related to choosing and administering assessments and significantly worse on communicating results. According to the authors, the results "give empirical evidence of the anticipated woefully low levels of assessment competency for teachers" (p.67). The results also showed that teachers who had had a measurement course performed better than teachers who lacked this background.

In summary, the small amount of existing literature on classroom assessment practices indicates that teachers probably need further training to improve the quality of the assessments that are used. There continues to be reliance on selected-response tests, with conflicting evidence concerning the use of essays. Whatever the type of question, few are written to tap students' higher level thinking skills. Appropriately, teachers appear to use a variety of assessment methods. There is clearly a need for more research on classroom assessments. Classroom assessments consume significant amounts of time for both teachers and students, and have important consequences. Particularly absent in the literature are examination of relationships between classroom assessment practices and grading, how teachers use assessments to set standards, and how teachers make decisions about the assessments they use.

Teachers' grading practices have received far more attention in the literature than have assessment practices. This may be due to the salient and summative nature of grades to students and parents. Grades have important consequences and communicate student progress to parents.

A study by Stiggins, Frisbie, and Griswold (1989) set the stage for research on grading by providing an analysis of current grading practices as related to recommendations of measurement specialists and newly established Standards for Teacher Competence in Educational Assessment of Students (American Federation of Teachers, National Council on Measurement in Education,

National Education Association, 1990). In this study the authors interviewed and/or observed 15 teachers on 19 recommendations from the measurement literature. They found that teachers use a wide variety of approaches to grading, and that they wanted their grades to both fairly reflect student effort and achievement, as well as to motivate students. Contrary to recommended practice, it was found that teachers value student motivation and effort, and set different levels of expectation based on student ability.

Brookhart (1994) conducted a comprehensive review of literature on teachers' grading practices. Her review identified 19 studies completed since 1984. Seven studies investigated secondary school grading, 11 studies both elementary and secondary, and one study elementary teachers. Three general methods of study were identified: surveys in which teachers responded to questions concerning components included in grading, grade distributions, and attitudes toward grading issues; surveys in which teachers were asked to respond to grading scenarios, asking what they would do in various circumstances; and qualitative methods, including interviews, observation, and document analysis. Despite methodological and grade level differences, the findings from these studies are remarkably similar. This suggests that conclusions warranted from the research are generalizable. Taken together, Brookhart comes to the following conclusions:

- Teachers inform students of the components used in grading.
- Teachers try hard to be fair in grading.
- Measures of achievement, especially tests, are major contributors to grades.
- Student effort and ability are used widely as components of grades.
- Elementary teachers rely on more informal evidence and observation, while secondary teachers use paper and pencil achievement tests and other written evidence as major contributors.
- Teachers' grading practices vary considerably from one teacher to another, especially in perceived meaning and purpose of grades, and how nonachievement factors will be considered.

- Teachers' grading practices are not consistent with recommendations of measurement specialists, especially confounding effort with achievement.

In one study, Brookhart (1993) investigated the meaning teachers give to grades and the extent to which value judgments are used in assigning grades. The results indicated that low ability students who tried hard would be given a passing grade even if the numerical grade were failure, while working below ability level did not affect the numerical grade. That is, an average or above average student would get the grade earned, whereas a below average student gets a break if there is sufficient effort to justify it. Teachers were divided about how to factor in missing work. About half indicated that a zero should be given, even if that meant a failure for the semester. The remaining teachers would lower the grade but not to a failure. The teachers' written comments showed that they strived to be "fair" to students. Teachers also seemed to indicate that a grade was a form of payment to students for work completed. More comments indicated that grades were something students earned as compared to grades indicating academic achievement, as compensation for work completed. This suggests that teachers, either formally or informally, include conceptions of student effort in assigning grades. Because teachers are concerned with student motivation, self-esteem, and the social consequences of giving grades, using student achievement as the sole criteria for determining grades is rare. This is consistent with earlier work by Brookhart (1991), in which she pointed out that grading often consists of a "hodgepodge" of attitude, effort, and achievement.

Cross and Frary (1996) report similar findings concerning the "hodgepodge" nature of grades. They surveyed 310 middle and high school teachers of academic subjects in a single system. A teacher survey was used to describe grading practices and opinions regarding assessment and grading. Consistent with Brookhart, it was reported that 72% of the teachers

raised the grades of low ability students. One-fourth of the teachers indicated that they raise grades for high effort "fairly often." Almost 40% of the teachers indicated that student conduct and attitude were taken into consideration when assigning grades. Interestingly, a very high percentage of teachers agreed that effort and conduct should be reported separately from achievement. Over half of the teachers reported that class participation was rated as having a moderate or strong influence on grades.

An earlier statewide study by Frary, Cross, and Weber (1993), using the same teacher survey that was used by Cross and Frary (1996), found similar results. Percentages of teachers agreeing or tending to agree to the following statements illustrate this conclusion:

Item	Percentage
• A student's ability should be taken into consideration in awarding final grades.	66
• An exceptionally low or high degree of student effort should be recognized by adjustment of the final grade.	66
• The amount of knowledge a student *gains* over the instructional period should be taken into consideration in awarding the final grade.	85
• Laudatory or disruptive classroom behavior should be considered in determining final grades.	31
• The minimum passing score on a test should be based at least in part on the scores earned by students of marginal ability who have be been putting forth satisfactory effort.	64

Another recent study by Truog and Friedman (1996), further confirms the notion of hodgepodge grading. In their study the written grading policies of 53 high school teachers were analyzed in relation to grading practices recommended by measurement specialists, and a focus group of eight teachers was conducted to probe reasoning used by the teachers. The study was based on an earlier investigation by Stiggins, Frisbie, and Griswold (1989). Friedman and Manley (1991) also found that teachers routinely use ability, attitude, effort, participation, and

other factors in addition to achievement when determining grades. Truog and Frieman (1996) found that written policies were consistent with earlier studies of teacher beliefs and practice. Nine percent of the teachers included ability as a factor in determining grades, 17% included attitude, 9% included effort, 43% included attendance, and 32% included student behavior.

Another survey of 143 elementary and secondary school teachers conducted by Cizek, Fitzgerald and Rachor (1995) collected data on teachers' assessment-related practices. Results indicated that assessment practices "were highly variable and unpredictable from characteristics such as practice setting, gender, years of experience, grade level or familiarity with assessment policies in their school district" (p. 159). Furthermore, teachers generally use a variety of objective and subjective factors to maximize the likelihood that students obtain good grades. Overall, the authors concluded that "many teachers seemed to have individual assessment policies that reflected their own individualistic values and beliefs about teaching" (p.160). The authors argue that grades should be used in more meaningful ways to communicate about student performance.

In summary, the literature on grading strongly supports the notion that teachers believe it is important to combine nonachievement factors, such as effort, ability, and conduct, with student achievement, to determine grades. While the studies are clear in this conclusion, less is known about how teachers decide to weigh these nonachievement factors in determining grades. Also, many of the surveys and other approaches in previous studies have asked teachers about their beliefs or projected behavior based on scenarios. It is possible that actual grading practice may be different. Despite increased focus on assessment and teacher competence with respect to measurement and grading, there appears to be a continuing discrepancy between recommended practice and teacher beliefs about grading. Furthermore, while descriptions of grading practices

are plentiful, there is little research on the relationship between grading practices and student motivation and achievement.

The literature reviewed on the nature and effect of assessment and grading practices on student achievement has demonstrated that there is little empirical evidence of the specific effects of using particular assessments and grading procedures. This is due in part to the complex nature of teaching, and how assessment and grading are only a part of instruction. Assessment and grading continue to be a private activity, with considerable variation among teachers. While "newer" forms of assessment, such as performance-based and portfolio, are based on recent research on cognitive learning, the suggestions are based on theory and not empirical evidence. There are several studies which show that teachers engage in assessment and grading practices that are not consistent with what would be recommended by measurement "experts." For example, combining nonachievement factors like effort, ability, and conduct with student achievement to determine grades, as well as "hodgepodge" grading. While descriptions of grading practices are plentiful, there is little research on the relationship between grading practices and student motivation and achievement. One theoretical model postulated by Brookhart (1997) represents an initial perspective about how assessment and grading practices affect self-efficacy, effort, and achievement. There is a strong research base with respect to the two major contributors to motivation (self-efficacy and importance, utility, and value), but not much about how specific assessment and grading practices effect these two components.

Research Questions

The purpose of the proposed research is to gather information from teachers regarding their assessment and grading practices to answer the following questions:

- What is the current state of assessment practice and grading by teachers?
- What assessment and grading topics do teachers identify as needs to be addressed in in-service?
- What is the relationship between assessment and grading practices and grades given to students?
- What are the relationships between grade level, and subject taught and assessment and grading practices?
- What are the reasons teachers give for their assessment and grading decision-making?
- What is the impact of the SOL tests on the extent to which different assessment techniques are used in the classroom?
- What classroom assessment and in-service needs do teachers have?

Methodology

Research Design

The research consisted of two phases, one involving a written survey of a large number of teachers and one using face to face interviews. Phase 1 included development and administration of a teacher questionnaire to survey teachers' assessment and grading practices and in-service needs. Quantitative analysis of the data included data reduction, descriptive statistical results, and the investigation of relationships with analysis of variance and correlational procedures. Phase 2 used interviews with selected teachers to investigate decision-making and justification for specific assessment and grading practices.

This report is concerned with Phase 1 of the research. A nonexperimental survey research design was utilized.

Population and Sample

The population included the entire population of grade 3-5 regular elementary teachers and all middle and high school science, mathematics, social studies, and English teachers in the seven school districts that are members of MERC. These divisions represent the entire metropolitan Richmond area. The return rate for all grade levels combined was 62% . A summary of the final sample is provided in Table 1.

Table 1

Summary of Final Sample and Return Rate

	Elementary	Middle	High School	Total
Population	1397	1105	1188	3690
Sample	921	633	850	2404
Return Rate	65%	54%	72%	65%

Instrument

The questionnaires were initially developed by the principal investigator early in 1997. The purpose of each questionnaire (one for elementary, one for secondary) was to document, using closed-form items, the extent to which teachers emphasized different assessment and grading practices, as well as in-service needs. A six point scale, ranging from *not at all* to *completely*, was constructed to allow teachers to indicate usage without the constraints of an ipsative scale that is commonly used in this area (e.g., percentage each factor contributes to grades). Also, the questions were worded to emphasize actual teacher behaviors in relation to a

specific class of students, rather than more global teacher beliefs. Separate questionnaires were developed for elementary and secondary levels (see Appendix A for copies of the questionnaires). At the secondary level, teachers were asked to identify a single class taught first semester and then answer all questions with this class in mind. At the elementary level, teachers responded to all items once for language arts and once for mathematics. The stem for the items was:

To what extent were final first semester grades of students in your single class described above based on:

The initial set of items was drawn from previous questionnaires that had been reported in the literature, as well as research on teachers' assessment and grading practices (Frary, Cross & Weber, 1993; Stiggins & Conklin, 1992; Brookhart, 1994). The items included factors that teachers consider in giving grades, such as student effort, improvement, academic performance, types of assessments used, and the cognitive level of the assessments (e.g., knowledge, application, reasoning). Additional items were added to measure grade level, and, for the secondary questionnaire, content area (mathematics, science, English, and history/social science) and ability level taught (advanced placement or honors, standard, or basic). Content-related evidence for validity for the initial draft of 47 items was strengthened by asking 42 classroom teachers (15 elementary, 12 middle, and 15 high school) to review the items for clarity and completeness of covering most if not all assessment and grading practices used. Appropriate revisions were made to the items, and a second pilot test with a school division outside of the MERC consortium was used to gather additional feedback on clarity, relationships among items, item response distributions, and reliability. Teachers from eight schools participated in the second pilot test, including 23 elementary, 26 middle, and 36 high school teachers. Item

statistics were used to reduce the number of items to 27. Items that showed a high correlation or minimum variation were eliminated, as well as items that were weak in reliability. Reliability was assessed by asking 28 of the teachers in the second pilot test to retake the questionnaire following a four week interval. The stability estimate was done by examining the percentage of matches for the items. Items that showed an exact match of less than 60% were deleted or combined with other items. The revised questionnaire included 34 items in the three categories (19 items assessing different factors used to determine grades, 11 items assessing different types of assessments used, and 4 items assessing the cognitive level of the assessments). The average exact match for the items was 46% of the teachers; 89% of the matches were within one point on the six point scale. Additional items asked teachers to indicate the approximate grade distribution of the class and the importance of assessment and grading topics for in-services.

Procedure

Three of the seven participating MERC school divisions administered the questionnaire in the spring of 1997; the remaining four MERC school divisions administered the questionnaire in February of 1998, soon after the end of the first semester. School division central administrators communicated to teachers that the questionnaire was to be completed, and were responsible for distribution and collection. The questionnaire took about 15 minutes to complete. Teachers were assured that their responses would be confidential. No information was on the form that could be used to identify the teacher. Teachers were given the opportunity to write their names on the questionnaire if they were interested in participating in a follow-up interview.

Data Analyses

The data analyses were primarily descriptive, using frequencies, percentages, means, medians, standard deviations, and graphic presentations to summarize overall findings and trends. An exploratory factor analysis was used to reduce the number of components investigated within each of the three categories of items. Relationships between assessment and grading practices, grades given, grade level, and subjects, were examined through multiple regression and analysis of variance procedures.

Findings

The findings are presented separately for elementary and secondary levels. The descriptive results are presented first, followed by relationships. The assessment and grading practices reported are organized by the three categories of items: factors used in grading, types of assessments used, and cognitive level of assessments.

Elementary

A total of 921 elementary teachers completed questionnaires. Of that number, 34% were at the third grade, 30% were at the fourth grade, 23% were at the fifth grade, and 17% were in classes with combined grades.

Descriptive Results The means and standard deviations for the three assessment and grading practices categories for both language arts and mathematics are reported in Table 2, grades given in Table 3, and in-service needs in Table 4. Table 5 shows the frequency distributions of a few questions to illustrate the spread of responses across the different points in the scale.

Table 2

Means and Standard Deviations of All Items Measuring Assessment and Grading Practices for Elementary Teachers
(n=873)

	Mathematics		Language Arts	
	Mean	SD	Mean	SD
Factors Used in Determining Grades				
1. Disruptive student behavior	1.37	.77	1.38	.77
2. Improvement of performance since the beginning of the year	3.00	1.20	3.07	1.21
3. Student effort-how much the student tried to learn	3.21	1.03	3.26	1.02
4. Ability levels of the students	3.39	1.31	3.40	1.29
5. Work habits and neatness	2.68	1.05	2.80	1.05
6. Grade distributions of other teachers	1.35	.85	1.33	.81
7. Completion of homework (not graded)	2.80	.98	2.77	.99
8. Quality of completed homework	2.69	1.13	2.73	1.14
9. Academic performance as opposed to other factors	4.40	1.06	4.37	1.07
10. Performance compared to other students in the class	2.00	1.03	2.04	1.03
11. Performance compared to a set scale of percentage correct (e.g., 86 -94%)	4.68	1.03	4.50	1.08
12. Performance compared to students from previous years	1.29	.71	1.31	.73
13. Specific learning objectives mastered	4.53	.92	4.50	1.08
14. Formal or informal school or district policy of the percentage of students who may obtain As, Bs, Cs, Ds, Fs	1.50	1.15	1.50	1.14
15. The degree to which the student pays attention and/or participates in class	3.01	1.07	3.10	1.07
16. Inclusion of 0s for incomplete assignments in the determination of final percentage correct	3.04	1.27	3.07	1.24
17. Extra credit for non academic performance (e.g., bringing in items for food drive)	1.34	.75	1.35	.77
18. Extra credit for academic performance	2.57	1.10	2.56	1.10
19. Effort, improvement, behavior and other "nontest" indicators for borderline cases	2.99	1.01	3.00	1.00
Types of Assessments Used in Determining Grades				
1. Major exams	3.21	1.39	3.05	1.38

2. Oral presentations	2.37	1.11	3.03	.88
3. Objective assessments (e.g., multiple choice, matching, short answer)	3.82	1.07	3.75	1.01
4. Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	2.84	1.14	3.43	.93
5. Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	3.54	1.05	3.22	1.06
6. Assessments designed primarily by yourself	3.63	.95	3.90	.98
7. Essay-type questions	2.42	1.15	3.39	1.03
8. Projects completed by teams of students	2.51	1.03	2.91	.99
9. Projects completed by individual students	3.06	1.24	3.59	.96
10. Performance on quizzes	3.93	.91	3.80	.98
11. Authentic assessments (e.g., "real world" performance tasks	2.95	1.08	2.89	1.06
Cognitive Level of Assessments Used in Determining Grades				
1. Assessments that measure student recall knowledge	3.65	.90	3.52	.86
2. Assessments that measure student understanding	4.46	.78	4.46	.77
3. Assessments that measure how well students apply what they learn	4.31	.84	4.28	.82
4. Assessments that measure student reasoning (higher order thinking)	3.99	.87	4.03	.86

The means and standard deviations in Table 2 show that, for this group of teachers as a whole, there are a few factors that contribute very little, if anything, to grades, namely:

- disruptive student behavior
- grade distributions of other teachers
- performance compared to other students
- school division policy about the percentage of students who may obtain different grades
- extra credit for nonacademic performance

Also, a few factors clearly contribute most, ranging from "quite a bit" to "extensively":

- academic performance as opposed to other factors
- performance compared to a set scale of percentage correct
- specific learning objectives mastered

The remaining factors contribute some:

- improvement of performance
- student effort
- ability levels of students
- work habits and neatness
- completion of homework
- quality of completed homework
- class participation and attention
- inclusion of zeros in calculating grades
- effort, improvement, behavior and other "nontest" indicators for borderline cases

There is a fairly large standard deviation reported for these items, showing considerable variation in the extent to which the factors are used for grading. For instance, the mean for student effort is 3.19 (some), with a standard deviation of 1.02. This suggests that at least 14% of the teachers responded "not at all" or "very little", and another 14%, at least, responded "quite a bit," "extensively," or "completely." In fact, as shown in Table 5, the percentages are 20% and 36%, respectively. This is a large percentage of teachers using effort in vastly different ways for grading. This same kind of dispersion of scores is evident in many of the factors. For example,

13% of elementary teachers reported using improvement "not at all" while 30% of the teachers responded "quite a bit," "extensively," or "completely." The extent to which ability level is used also shows great variability, with 23% of the teachers responding "not at all" and 47% responding "quite a bit," "extensively," or "completely." As we will see with other data, this pattern of high variability is one of the major findings of the research.

Given that the grading scales in the divisions used in the study are based on how performance compares to a set scale of percentage correct (e.g., 94-100 A, 86-93 B, and so on), it was surprising to find that only 65 percent of the teachers responded that they used this "extensively" or "completely."

The items that asked teachers about the types of assessments used shows that teachers do not rely on a single kind of assessment. Rather, many different types of assessments appear to be utilized. While objective assessments are used most frequently, performance assessments and projects are used almost as much in language arts (means of 3.75, 3.43, and 3.59, respectively). There is great reliance on assessments prepared by the teachers themselves, but also considerable use of assessments provided by publishers (means of 3.90 and 3.22, respectively). The lowest rated type of assessment, in terms of use, for language arts, was authentic assessments (mean of 2.89). This suggests some use of authentic assessment for most teachers. Interestingly, the mean for performance assessments in math was slightly higher (2.95). The standard deviations with respect to types of assessments (about 1 point on the scale) point to considerable variation.

Cognitive levels of assessments were very similar for math and language arts. The lowest rated assessments, in terms of use, were those that measure student recall knowledge. The highest was student understanding, with application and reasoning in between. For the three highest rated items the means were around 4 on the scale (used Quite a Bit).

Grades The results for percentages of different grades awarded by elementary teachers are presented in Table 3. The table is broken out by grade level and subject matter as well as letter grade awarded. Percentages are estimated by teachers and therefore may not sum to 100%.

Grades of A, B and C are most typically awarded by elementary teachers, comprising more than 70% of the total grades given. Grades of D and F comprise less than 10% of total grades given. A grade of B is most typically awarded by teachers, accounting for approximately 32 to 35 % of total grades given. Grades of A and C are nearly equally distributed, accounting for approximately 40% of a combined total, with grades of A comprising approximately 18 to 24% of total grades given and grades of C comprising approximately 21 to 25%. Grades of D are awarded between 6 and 8% of the total, while grades of F are awarded less than 3% of the total grades given.

Table 3

Percentages of Different Grades Awarded by Elementary Teachers
(n=859)

	A		B		C		D		F	
Grade Level	Math	LA	Math	LA	Math	LA	Math	LA	Math	LA
3 (n=294)	23.97	22.56	35.02	34.20	21.33	23.77	6.29	6.24	2.24	1.78
4 (n=258)	21.39	22.27	35.50	34.73	24.52	24.35	6.67	6.22	2.90	2.33
5 (n=205)	17.56	20.83	33.32	32.24	23.60	23.90	7.54	7.31	2.62	2.33
Mixed (n=102)	19.56	18.52	32.18	33.08	21.91	22.48	7.11	6.07	2.54	1.75
Total	20.62	21.05	34.00	33.56	22.84	23.63	6.90	6.46	2.58	2.05

In-Service Needs The results of teacher in-service needs are reported in Table 4. Nineteen needs were surveyed, and over half of them had a mean score above the midpoint of 3.5, thus indicating strong teacher in-service training needs in those areas. Of highest need to teachers is training in the assessment of reading proficiency, with a mean of 4.03. Understanding and using the new SOL (Standards of Learning) tests was the second highest need of teachers, with a mean of 3.92. The assessment of reasoning and other "higher order" thinking skills, along with the assessment of writing skills, share the third place position for most important needs of teachers, with means of 3.86 and 3.84 respectively.

Other high priority needs with means above 3.7 included items such as the improvement of overall quality of classroom assessments and communication with parents about grades and test scores.

Needs with means above 3.6 included the use of assessment information during instruction, understanding and using the new Stanford 9 standardized tests, using assessment results to evaluate instruction, and understanding the link between assessment and instruction.

The three lowest priority needs of teachers were identified as calculating final course grades, using portfolio assessments, and designing paper and pencil tests, with means of 2.82, 2.86 and 2.97 respectively.

Table 4

Means and Standard Deviations of In-Service Needs Indicated by Elementary Teachers
(n=102)

Item	Mean	SD
1. Using assessment information for planning prior to instruction	3.26	1.00
2. Using assessment information during instruction (e.g., monitoring student progress, judging whether students understand, questioning students)	3.65	1.05
3. Using assessment results to evaluate instruction and curriculum	3.60	.99
4. Using assessment results to determine student grades	3.53	1.03
5. Communicating with parents concerning grades and test scores	3.70	1.07
6. Understanding and using the new Stanford 9 standardized tests	3.64	1.05
7. Understanding and using the new SOL tests	3.92	.98
8. Understanding technical assessment concepts such as reliability and validity	3.09	1.06
9. Improving the overall quality of classroom assessments	3.76	.93
10. Assessing reasoning and other "higher order" thinking skills	3.86	.91
11. Using performance-based assessments, such as presentations and projects	3.44	.93
12. Using portfolio assessments	2.86	1.00
13. Designing paper and pencil tests (e.g., multiple choice, short answer, essay)	2.97	1.03
14. Assessing writing skills	3.84	.95
15. Assessing reading proficiency	4.03	.93
16. Assessing mainstreamed students	3.55	1.03
17. Assessing affective traits, such as attitudes, values and self-concept	3.18	1.03
18. Understanding the link between assessment and instruction	3.60	1.08
19. Calculating final course or semester grades	2.82	1.26

Table 5

Percentages of Elementary Teachers' Responses to Selected Items for Mathematics Assessment Practices and Grading

Factors Contributing to Grades

Question	Not at All	Very Little	Some	Quite a Bit	Extensively	Completely
Improvement of performance since the beginning of the year	13	17	38	21	7	2
Student effort – how much the student tried to learn	6	14	44	27	7	2
Ability levels of the students	10	13	31	24	19	4
Academic performance compared to other factors	2	3	12	29	44	10
Performance compared to other students in the class	40	30	21	7	2	0
Performance compared to a set scale of percentage correct	1	1	10	22	45	20

Types of Assessments Used

Question	Not at All	Very Little	Some	Quite a Bit	Extensively	Completely
Objective assessments	2	8	28	36	21	5
Performance Assessments	14	23	38	17	7	1
Assessments designed primarily by yourself	1	7	43	30	17	2
Authentic assessments	10	20	44	18	8	1

Cognitive Level of Assessments

Question	Not at All	Very Little	Some	Quite a Bit	Extensively	Completely
Assessments that measure student reasoning	0	2	25	44	25	4

The relatively large variability of teacher responses is illustrated by the standard deviations. Table 6 looks at another procedure to examine variability by comparing variability within schools to variability between schools. To calculate the average standard deviation within schools, the responses of teachers from the same school were used to derive a standard deviation score for that school for each item. A total of 105 standard deviations, one for each of 105 schools, were then averaged to result in within school variability. Between school variability was calculated by using the mean for each school, considering that as a single score, and then calculating the standard deviation of the means. The results of these analyses for three items, and percentage of As awarded, are summarized in Table 6.

Table 6

Variation Within and Between Elementary Schools for Selected Items
(n=105 schools and teachers)

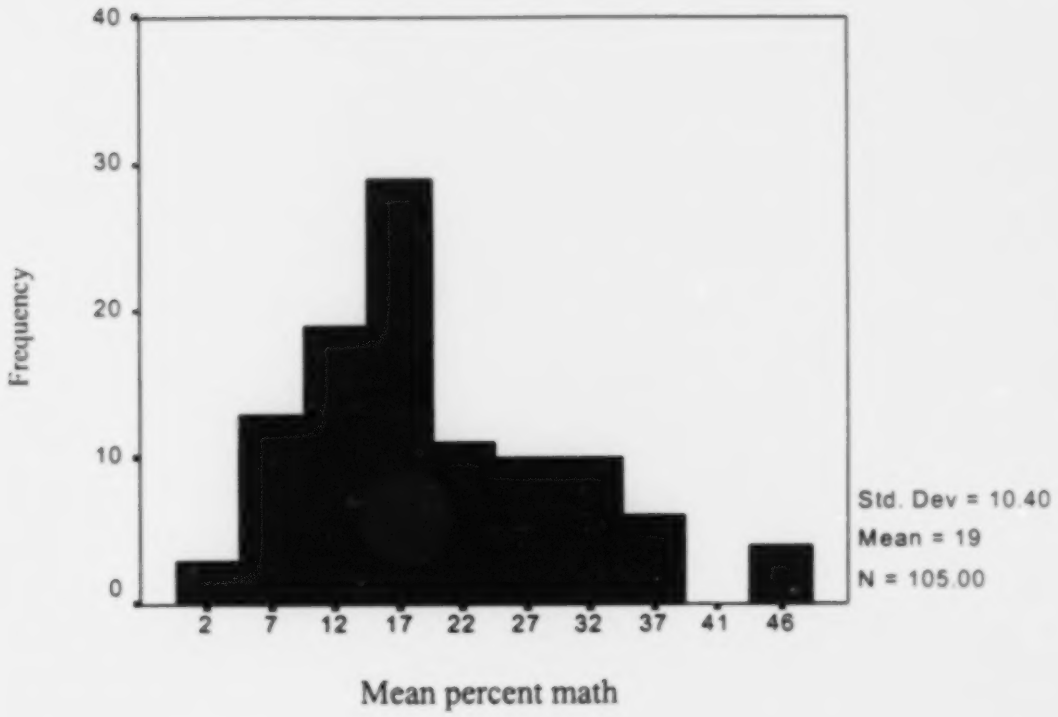
Question	Mean Variation Within	Mean Variation Between
% As awarded in math	16.2	10.4
Student effort – how much the student tried to learn	.92	.57
Assessments that measure student reasoning	.81	.42
Objective assessments	.97	.51

In each case the average variation within schools is greater than the variation between schools. While this result is influenced by the relatively low number of teachers in each school, which would increase the variation, it still suggests that teachers in the same school differ more, on the average, than responses compared at the school level.

The chart in Figure 1 illustrates the frequency of mean percentage math As awarded between schools.

Figure 1

Between School Variability and As Awarded for Elementary Teachers



It shows that the percentage of math As awarded was 12 percent or less for thirty five elementary schools, while for 20 schools the percentage of math As awarded was 32 percent. This shows a large between school variation of the number of As awarded.

Data Reduction Prior to examining the relationships between subject (mathematics compared to language arts) and grade level (grades 3, 4, and 5), a data reduction process was performed for each of the major categories of items (factors, types, and cognitive levels) for both mathematics and language arts. The first step in the data reduction was to eliminate items that showed a floor effect with little variability. The remaining items were used in the second step of the data reduction, a factor analysis, to identify relationships among the items by reducing them to a few relatively independent but conceptually meaningful composite variables called components. A varimax rotation was used for the factor analyses.

The factor analysis for items used in grading (factors) resulted in six components. There were no differences between mathematics and language arts. The loadings of different items are summarized in Table 7.

Table 7

Factor Loadings for Elementary Teachers' Assessment and Grading Practices

Math Factors	Factor Loading
Component 1	
Improvement of performance since the beginning of the year	.783
Student effort - how much the student tried to learn	.819
Ability levels of the students	.629
Component 2	
Completion of homework (not graded)	.824
Quality of completed homework	.724
Component 3	
Grade distributions of other teachers	.660
Component 4	
Academic performance as opposed to other factors	.716
Performance compared to a set scale of percentage correct (e.g., 86-94% B)	.659
Specific learning objectives mastered	.699
Component 5	
Performance compared to other students in the class	.772
Performance compared to students from previous years	.737
Component 6	
Extra credit for academic performance	.732
Effort, improvement, behavior and other "nontest" indicators for borderline cases	.652
Math Types	
Component 1	
Oral Presentations	.614
Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	.786
Essay-type questions	.740
Projects completed by teams of students	.819
Projects completed by individual students	.712
Authentic assessments (e.g., "real world" performance tasks)	.616
Component 2	
Objective assessments (e.g., multiple choice, matching, short answer)	.691
Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	.832

Component 3	
Major exams	.691
Assessments designed primarily by yourself	.611
Math Cognitive Abilities	
Component 1	
Assessments that measure student understanding	.791
Assessments that measure student reasoning (higher order thinking)	.786
Assessments that measure how well students apply what they learn	.820
Language Arts Factors	
Component 1	
Improvement of performance since the beginning of the year	.771
Student effort - how much the student tried to learn	.795
Ability levels of the students	.654
Component 2	
Completion of homework (not graded)	.809
Quality of completed homework	.777
Component 3	
Disruptive student behavior	.622
Grade distributions of other teachers	.661
Component 4	
Academic performance as opposed to other factors	.697
Performance compared to a set scale of percentage correct (e.g., 86-94% B)	.682
Specific learning objectives mastered	.674
Component 5	
Extra credit for academic performance	.729
Effort, improvement, behavior and other "nontest" indicators for borderline cases	.665
Component 6	
Performance compared to other students in the class	.790
Performance compared to other students in the class	.745
Language Arts Types	
Component 1	
Oral presentations	.755
Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	.706
Projects completed by teams of students	.711

Authentic assessments (e.g., "real world" performance tasks)	.670
Component 2	
Objective assessments (e.g., multiple choice, matching, short answer)	.796
Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	.706
Performance on quizzes	.736
Component 3	
Assessments designed primarily by yourself	.840
Essay-type questions	.672
Language Arts Cognitive Abilities	
Component 1	
Assessments that measure student understanding	.806
Assessments that measure student reasoning (higher order thinking)	.815
Assessments that measure how well students apply what they learn	.828

The first component was comprised by three items that emphasized effort, ability, improvement, work habits, attention, and participation. These items could be considered enablers to academic performance, important indicators to teachers to judge the degree to which the student has tried to learn, and, by implication, actually learned. A second component was defined by the two items that asked about homework. The third component included three items that focused on academic performance of the student. The fourth component loaded on one item concerning extra credit and one for borderline cases. Thus, there appears to be four conceptually meaningful variables that teachers use when grading students: actual performance, effort and improvement, homework, and borderline cases. Given the relatively low emphasis on homework and the infrequent occurrence of borderline cases, these results suggest that teachers conceptualize two major ingredients: actual performance, and effort and improvement. Of these two, clearly academic performance is more important.

The factor analysis for types of assessments used resulted in three components for both mathematics and language arts. The item loadings were, for the most part, the same for both subjects. The first component was comprised by six items for math types and four items for language arts types, each of which described some kind of constructed-response assessment, such as essays (math only), projects, and performance assessments. The second component, made up of either two or three items, included objective assessments, quizzes (language arts only), and assessments provided by publishers. Evidently items provided by publishers are used in both quizzes and objective assessments. The third component was comprised of two items for math (major exams and teacher-made tests) and two items for language arts (teacher-made tests and essays). This suggests that the common element in the third component is "teacher-made." For math the major exams tend to be teacher-made and for language arts essays tend to be teacher made.

The factor analysis for cognitive levels showed high intercorrelation among the three items that suggested "higher order" knowledge and skills (understanding, reasoning, and application). Teachers tend to think about these as one kind of skill, apart from recall knowledge, which did not load on this analysis.

Relationship Results The relationship analyses for subject matter and grade level used *t*-tests and analysis of variance using standardized component scores for the items loading on each of the eight components derived from the factor analyses, plus the percentage of As given, as dependent variables. A regression analysis was also performed to determine if assessment and grading practices predict grades. Thus, there were two independent variables, subject matter,

with two levels, and grade level, with three levels, in the first two analyses; all eight components were used as independent variables to predict the percentage of As awarded.

Paired *t*-tests were used to identify any differences between math and language arts responses. The *t*-test analyses showed that there are few differences between language arts and mathematics assessment and grading practices, despite the large sample size that would make it easy to detect significant differences. Clearly there is more in common than there is a difference on the basis of these two content areas. As might be expected, differences occurred for the extent to which performance assessments were used (mean of 2.33 for math and 3.41 for language arts), projects completed by individual students (mean of 3.01 for math and 3.56 for language arts), and the use of assessments provided by publishers (mean of 3.56 for math and 3.23 for language arts). Thus, only three items in the category types of assessments used showed a difference between math and language arts. When considering other factors such as effort, participation, homework, etc., as well as cognitive levels, there was no difference between the math and language arts responses.

One-way analysis of variance analyses, with Sheffe' post hoc tests, were used to examine the relationship between grade level and assessment and grading practices. The results of these analyses are shown in Table 8, which summarizes the components that indicate no relationship, those that show a positive relationship, and the single variable that showed a negative relationship.

Table 8

Relationship of Grade Level (3,4, and 5) to Assessment Practices of Elementary Teachers
(n= 873)
Math

No Relationship	Positive Relationship	Negative Relationship
Effort, ability, & improvement	Homework	Percent As
Academic performance	Extra Credit	
Teacher-made major exams	Constructed-Response Assessments	
Higher order thinking & application	Objective Assessments	

Language Arts

No Relationship	Positive Relationship	Negative Relationship
Effort, ability, & improvement	Homework	
Academic performance	Extra Credit	
Objective Assessments	Constructed-Response Assessments	
Higher order thinking & application	Teacher-made major exams	
Percent As		

As with other analyses, the major finding is no difference between grade levels on components that are most important to assessment and grading. For both language arts and mathematics the results show that as grade level increases so does the importance of homework, extra credit, and constructed-response assessments. For math, the importance of objective assessments shows a positive relationship with grade level. In language arts teacher made major exams contribute more in higher grades. The only negative relationship was found in the percentage of As awarded, which means that fewer As are awarded in higher grades.

The predictive relationship between assessment and grading practices was examined with stepwise multiple regression, one for language arts, one for mathematics, with percentage As

awarded as the dependent variable and the eight weighted component scores as independent variables. The results of these regressions are summarized in Table 9.

Table 9

Factors, Types of Assessments, and Cognitive Levels as Predictor Variables of Percent As Awarded for Elementary Teachers

	R	Significant Positive Relationship	Significant Negative Relationship
Mathematics (n=714)	.21	Higher order thinking and application	Objective assessments Publisher-provided items Homework
Language Arts (n=731)	.20	Constructed-response items	Extra credit

The multiple correlation coefficients are relatively small in both regressions, suggesting that the major predictors of grades are not the weight given to different factors, types of assessments, or cognitive level of assessments. Given that finding, the percentage of As awarded tended to increase with increased weight given to higher order thinking assessments math, and constructed-response assessments for language arts. Negative relationships for math were found with objective assessments, publisher-provided items, and homework for math, and extra credit for language arts.

Summary of Elementary Level Findings The results of the elementary level analyses are summarized as follows:

- Most teachers use a multitude of factors in grading students.
- Academic performance is clearly the most important factor in grading students but non-test performance and behavior, such as effort, participation, and extra credit work, are also very important.
- Disruptive student behavior, grade distributions of other teachers, and norm-referenced interpretations contribute little to grading.

- District or school grading policies related to the percentages of students who may obtain different grades contribute little to grading.
- A substantial number of teachers include zeros in the calculation of grades.
- There are four major components in the various factors teachers use for grading: academic performance; effort, improvement, and ability; homework; and extra credit.
- Three major types of assessments are used: constructed response, such as projects, essays, and presentations; objective assessments; and teacher-made major exams. While objective assessments are used most frequently, there is also a great reliance on constructed response types of assessments.
- There is a tendency for teachers to differentiate the cognitive level of their assessments into two categories: recall knowledge and "higher order" thinking and application. "Higher order" thinking and application are emphasized heavily.
- There is a significant reliance on assessments that are designed by publishers.
- There is great variation within schools concerning the extent to which teachers emphasize different factors in grading students.
- Greater emphasis is placed in later grades on homework, extra credit, constructed-response assessments, objective assessments, and major exams. Other practices, such as effort, ability, improvement, and academic performance are emphasized the same in all three grade levels.
- Teachers who award more As use fewer objective assessments, fewer publisher-provided tests, less homework, and more assessments that measure reasoning and application. There was no relationship between the extent to which effort, improvement, ability, academic performance, homework and extra credit were emphasized, and As awarded.

Secondary

The secondary teachers (middle and high school) were asked on the survey to answer all questions with a single class in mind, the class that they taught most frequently. This was done to provide a more specific point of reference for the teachers that would clarify interpretation of the data. Otherwise, responses would blend practices used in several different types of classes.

Table 10 shows the number of classes broken out by subject, grade level, and ability level of the class for both middle and high school teachers. Further interpretations of the findings need to keep this distribution in mind.

Table 10
Number of Secondary Teachers by Subject, Grade Level and Ability Level

		Middle School		
Grade Level	**6**	**7**	**8**	
	182 (35%)	168 (33%)	165 (32%)	
Subject	**Math**	**English**	**Science**	**Social Science**
	140 (27%)	154 (30%)	115 (23%)	94 (18%)
Ability Level	**Honors**	**Standard**	**Basic**	**Mixed**
	116 (23%)	213 (42%)	57 (11%)	112 (22%)

		High School		
Grade Level	**9**	**10**	**11**	**12**
	220 (39%)	136 (24%)	124 (22%)	90 (16%)
Subject	**Math**	**English**	**Science**	**Social Science**
	80 (14%)	196 (35%)	146 (26%)	144 (25%)
Ability Level	**AP/Honors**	**Standard**	**Basic**	**Mixed**
	134 (24%)	285 (51%)	75 (13%)	64 (11%)

Descriptive Results The means and standard deviations for the three assessment and grading practices categories for middle and high school teachers are reported in Table 11. The in-service needs are shown in Table 12, and the percentages of grades given to students is summarized in Table 13. Table 14 shows the frequency distribution of a few questions to illustrate the variability of scores across different values of the scale.

An examination of the means of the assessment and grading factors reported in Table 11 indicates very few differences between middle and high school teachers. There is some indication that middle school teachers tend to use student effort and assessments provided by publishers more than high school teachers, and that high school teachers use more major exams, more comparisons with other students, and emphasize 0s in grading more than middle school teachers. Otherwise, there is little difference between middle and high school teachers' assessment and grading practices.

Like elementary teachers, there appear to be a few items that contribute little or nothing to grading, including the following:

- Disruptive student behavior
- Grade distributions of other teachers
- Performance compared to students from previous years
- School division policy about the percentage of students who may obtain different grades
- Extra credit for nonacademic performance

There are also a few factors that clearly contribute the most to grading, with means at or above "quite a bit (4):"

- Academic performance as opposed to other factors
- Performance compared to a set scale of percentage correct
- Specific learning objectives mastered

Also very similar to elementary teachers, there are a number of factors that appear to contribute "some," (means at or above 3)

- Student effort
- Ability levels of students
- Quality of homework completed
- Class participation and attendance
- Inclusion of 0s

Elementary teachers tended to value completion of homework and work habits and neatness more than secondary teachers, though the remainder of the factors that contribute significantly to grading is virtually the same as secondary. Also like elementary, the large standard deviations shows considerable variation. This means that a large percentage of secondary teachers use many of these five factors to a great extent in determining grades. For example, the mean for student effort was 3.23, with a standard deviation of 1.11. By examining the frequency distribution for this question in Table 15, approximately 40% of the teachers responded "quite a bit," "extensively," or "completely." About 20% of the teachers indicated "not at all" or "very little" to using student effort. This represents a considerable difference among teachers in the extent to which effort is included in grading. This same kind of variation

occurs with other items that tend to average in the middle of the scale. As with elementary teachers, then, this large variation in practice is one of major findings of the study.

With respect to how grades are determined, it is surprising but consistent with elementary teachers, that only 55% responded that they use performance compared to a set scale of percentage correct "extensively" or "completely." Evidently, unless teachers misunderstood the question, there are many other determinants of grades than use of the set scale.

Concerning types of assessments used, there is great reliance on assessments designed primarily by the teachers themselves, with relatively little reliance on those provided by publishers. Essay type questions are used only slightly less than objective tests (means of 3.28 and 3.64, respectively), and there is considerable use of student projects and performance assessment by teachers (mean of 3.17; approximately 35% of the teachers use student projects and performance assessment at least "quite a bit"). However performance assessments appear to be used less by secondary teachers than elementary teachers. Oral presentations and authentic assessments are used least. There were very few differences between middle and high school teachers.

The cognitive levels of the assessments used were the same for middle and high school teachers. Student understanding was rated highest, with a strong emphasis on both reasoning and application. Recall knowledge was used least. These results match what was found for elementary teachers. It is interesting to note that a high percentage of teachers indicated that they use assessments measuring recall knowledge quite a bit (34%), extensively (11%), or completely (1%). While the percentages for measuring student understanding were higher (47%, 35%, 3%, respectively), it appears that for many teachers there nearly as much emphasis at the recall level as at understanding.

Table 11

Means and Standard Deviations of All Items Measuring Assessment and Grading Practices for Secondary Teachers

Item	Middle (N=630)		High (N=846)		Total (N=1506)	
Factors	Mean	SD	Mean	SD	Mean	SD
Disruptive student performance	1.5	.83	1.60	.91	1.56	.88
Improve of performance since the beginning of the year	2.86	1.14	2.83	1.12	2.85	1.13
Student effort-how much the student tried to learn	3.31	1.13	3.16	1.10	3.23	1.11
Ability levels of the students	3.38	1.33	3.43	1.28	3.41	1.30
Work habits and neatness	2.80	1.07	2.68	1.06	2.73	1.07
Grade distributions of other teachers	1.20	.65	1.18	.61	1.19	.62
Completion of homework (not graded)	3.02	1.06	2.95	1.12	2.98	1.10
Quality of completed homework (graded)	3.18	1.15	3.22	1.14	3.20	1.15
Academic performance as opposed to other factors	4.37	1.08	4.34	1.09	4.35	1.08
Performance compared to other students in the class	2.06	1.13	2.23	1.18	2.16	1.17
Performance compared to a set scale of percentage correct	4.44	1.24	4.45	1.31	4.43	1.29
Performance compared to students from previous years	1.45	.91	1.47	.85	1.46	.87
Specific learning objectives mastered	4.38	.92	4.35	.91	4.37	.92
Formal or informal school or district policy of the percentage of students who may obtain As, Bs, Cs, Ds, Fs	1.58	1.12	1.51	1.08	1.54	1.10
Degree to which the student pays attention and/or participates in class	3.12	1.11	3.20	1.12	3.17	1.12
Inclusion of 0s for incomplete assignments in the determination of final percentage correct	3.61	1.29	3.90	1.32	3.77	1.31
Extra credit for nonacademic performance (e.g., bringing in items for food drive)	1.54	.86	1.49	.76	1.51	.80
Extra credit for academic performance	2.66	1.18	2.54	1.06	2.60	1.11

Effort, improvement, behavior and other "nontest' indicators for borderline cases	2.91	1.11	2.82	1.08	2.87	1.09
Types of Assessments						
Major exams	2.53	1.29	3.43	.87	3.05	1.16
Oral presentations	2.52	1.04	2.40	1.03	2.46	1.04
Objective assessments (e.g., multiple choice, matching, short answer)	3.56	1.10	3.70	1.05	3.64	1.07
Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	3.19	1.17	3.14	1.20	3.17	1.19
Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	2.67	1.17	2.40	1.10	2.53	1.14
Assessments designed primarily by yourself	4.31	.99	4.55	1.03	4.44	1.02
Essay-type questions	3.17	1.16	3.37	1.26	3.28	1.22
Projects completed by teams of students	2.76	1.13	2.68	1.12	2.72	1.13
Projects completed by individual students	3.22	1.15	3.14	1.14	3.17	1.15
Performance quizzes	3.86	.86	3.75	.84	3.80	.85
Authentic assessments (e.g., "real world" performance tasks)	2.85	1.08	2.66	1.04	2.75	1.06
Cognitive Level of Assessments						
Assessments that measure student recall knowledge	3.48	.85	3.51	.86	3.50	.85
Assessments that measure student understanding	4.27	.77	4.25	.74	4.26	.76
Assessments that measure student reasoning (higher order thinking)	3.96	.90	4.02	.92	4.00	.91
Assessments that measure how well students apply what they learn	4.11	.88	4.08	.89	4.10	.89

In-Service Needs The results of secondary teachers' in-service needs are summarized in Table 12. Five of the items showed means greater than 3.5, indicating a fairly strong need:

- Using assessment during instruction
- Understanding and using SOL tests
- Improving overall quality of classroom tests
- Assessing reasoning and other higher order thinking
- Understanding the link between assessment and instruction

The two most important needs were assessing reasoning and other higher order thinking and improving the overall quality of classroom tests. This is consistent with the high use of teacher made assessments in the classroom and the relatively high emphasis of reasoning in the assessments. It is in contrast with elementary level teachers who rated assessment of reading and interpretation of the SOL tests highest. It is interesting to note that elementary teachers, overall, rate in-service needs higher than secondary teachers for all but a few of the areas.

Table 12

Means and Standard Deviations of In-Service Needs Indicated by Secondary Teachers

	Middle (N=633)		High (N=845)		Total (N=1507)	
	Mean	SD	Mean	SD	Mean	SD
Using assessment information during instruction (e.g., monitoring student progress, judging whether students understand, questioning students)	3.13	.97	2.93	1.00	3.02	.99
Using assessment information during instruction (e.g., monitoring student progress, judging whether students understand, questioning students)	3.62	1.02	3.42	1.02	3.52	1.02
Using assessment results to evaluate instruction and curriculum	3.53	.94	3.29	.99	3.40	.98
Using assessment results to determine student grades	3.38	1.05	3.21	1.13	3.28	1.10
Communicating with parents concerning grades and test scores	3.48	1.08	3.25	1.08	3.36	1.09
Understanding and using the new Stanford 9 standardized tests	3.18	1.15	2.83	1.19	3.00	1.19
Understanding and using the new SOL tests	3.71	1.06	3.34	1.15	3.51	1.13
Understanding technical assessment concepts such as reliability and validity	3.05	1.03	2.90	1.04	2.98	1.04
Improving the overall quality of classroom assessments	3.75	.94	3.56	.95	3.65	.95
Assessing reasoning and other "higher order" thinking skills	3.87	.90	3.77	.94	3.82	.92
Using performance-based assessments, such as presentations and projects	3.43	.99	3.25	.98	3.34	.99

Using portfolio assessments	2.88	1.07	2.69	1.08	2.78	1.08
Designing paper and pencil tests (e.g., multiple choice, short answer, essay)	2.86	1.09	2.89	1.12	2.88	1.11
Assessing writing skills	3.38	1.13	3.27	1.13	3.32	1.13
Assessing reading proficiency	3.44	1.16	3.24	1.14	3.34	1.15
Assessing mainstreamed students	3.47	.96	3.21	1.06	3.33	1.03
Assessing affective traits, such as attitudes, values, and self-concept	3.06	1.06	2.90	1.07	2.98	1.07
Understanding the link between assessment and instruction	3.61	1.03	3.42	1.04	3.51	1.04
Calculating final course or semester grades	2.90	1.26	2.62	1.29	2.75	1.28

Grades reported by secondary teachers, broken out by grade level, are presented in Table 13. The grades of A and B are awarded to approximately 50% of the students in middle school, 36% in grades 9-11, and 48% in grade 12. The percentage of students receiving failing grades increases significantly in 9th grade and declines during grades 10-12. Like elementary teachers, B grades are the most common awarded in grades 9 and 12, while C is most common in grades 10 and 11.

Table 13

Percentages of Grades Awarded by Secondary Schools

Grade Level

Grade	6	7	8	9	10	11	12
A	20.53	16.42	18.07	10.25	9.83	10.93	14.64
B	33.96	35.41	32.35	26.39	27.39	27.17	33.74
C	25.13	25.34	27.60	27.25	31.57	27.65	26.71
D	9.78	11.73	11.95	14.87	15.06	14.61	11.64
F	4.67	5.55	6.44	12.55	11.35	9.67	6.48

Table 14

Percentages of Secondary Teachers' Responses to Selected Items for Assessment Practices and Grading

Factors Contributing to Grades	Grade Level	Not at All	Very Little	Some	Quite a Bit	Extensively	Completel
Improvement of performance since the beginning of the year	Middle	14.86	19.17	38.82	20.13	6.07	.96
	High	15.57	17.72	41.80	18.56	5.39	.96
Student effort – how much the student tried to learn	Middle	7.48	12.42	37.58	29.62	10.03	2.87
	High	7.81	16.33	40.24	24.62	9.59	1.42
Ability levels of the students	Middle	12.40	10.47	28.02	28.82	15.94	4.35
	High	10.40	10.52	29.99	27.21	18.86	3.02
Academic performance compared to other factors	Middle	2.73	3.21	9.95	32.74	40.77	10.59
	High	2.57	2.94	13.59	29.50	41.37	10.04
Performance compared to other students in the class	Middle	41.40	26.27	21.02	7.80	3.03	.48
	High	34.09	29.43	20.57	11.24	4.19	.48
Performance compared to a set scale of percentage correct	Middle	3.53	3.04	14.58	23.40	36.06	19.39
	High	4.45	4.69	11.43	20.82	37.55	21.06

Types of Assessments Used							
Objective assessments	Middle	3.48	11.87	32.44	32.12	17.41	2.69
	High	1.65	11.07	28.98	35.22	20.49	2.59
Performance assessments	Middle	9.37	15.87	36.03	25.40	11.43	1.90
	High	9.20	20.52	33.49	22.17	13.21	1.42
Assessments designed primarily by yourself	Middle	.48	1.59	19.68	33.65	33.49	11.11
	High	.47	1.06	16.98	23.58	40.68	17.22
Authentic assessments	Middle	11.20	24.03	41.23	15.91	6.98	.65
	High	15.10	26.55	40.80	13.15	3.90	.49
Cognitive Level of Assessments							
Assessments that measure student reasoning	Middle	.00	2.69	21.87	41.52	29.79	4.12
	High	.12	2.36	25.24	37.85	31.01	3.42

Similar to the findings for elementary teachers, there is a large variation in secondary teachers' assessment practices and grading. To capture this variability, Table 15 shows the results of an analysis that compares variation within schools to variation between schools. The standard deviations for a select number of items in each school were averaged to represent variation within, while the standard deviation of the mean scores for all schools was used as a measure of variation between schools. As shown in the table, within school variation is greater than between school variation. This suggests more differences between teachers in the same school than exists between teachers at different schools.

Table 15
Variation Within and Between Secondary Schools for Selected Items
(N= 1513 teachers)

Question	Mean Variation Within	Mean Variation Between
%As awarded in math	7.95	7.00
Student effort-how much the student tried to learn	.99	.73
Assessments that measure student reasoning	.95	.54
Objective assessments	1.14	.78

Data Reduction Prior to examining the relationships between grade level, subject, and ability level of class with assessment and grading practices, factor analysis was used to reduce the items to fewer more meaningful components. These analyses were done separately for middle school and high school teachers. Initially, items that showed a floor effect were eliminated. A varimax rotation was used for factors used in grading, types of assessments, and cognitive levels of assessments. The results of the analyses are summarized in Table 16.

Table 16

Factor Loadings for Middle and High School Teachers' Assessment and Grading Practices

	Factor Loadings	
Factors	Middle	High
Component 1		
Improvement of performance since the beginning of the year	.748	.730
Student effort - how much the student tried to learn	.808	.777
Ability levels of the students	.655	.603
Degree to which the student pays attention and/or participates in class	.610	.607
Work habits and neatness		.618
Component 2		
Academic performance as opposed to other factors	.735	.739
Performance compared to a set scale of percentage correct	.708	.662
Specific learning objectives mastered	.722	.668
Component 3		
Completion of homework (not graded)	.687	
Inclusion of 0s for incomplete assignments in the determination of final percentage correct		.639
Component 4		
Extra credit for academic performance	.807	.826
Effort, improvement, behavior and other "nontest" indicators for borderline cases		.692
Types		
Component 1		
Oral Presentations	.764	.726
Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	.778	.787
Essay-type questions	.633	.749
Projects completed by teams of students	.750	.747
Projects completed by individual students	.709	.739
Component 2		
Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	.800	-.868
Assessments designed primarily by yourself	-.884	.845
Component 3		
Objective assessments	.612	.661
Performance quizzes	.868	.782

Component 4		
Major exams	.967	
Cognitive Abilities		
Component 1		
Assessments that measure student understanding	.865	.816
Assessments that measure student reasoning (higher order thinking)	.873	.911
Assessments that measure how well students apply what they learn	.846	.873
Component 2		
Assessments that measure student recall knowledge		.981

The loadings for the factors identified the same four components for middle and high school teachers, with minor differences, and these four components were also found for elementary teachers. The four components included academic performance, academic-enabling behaviors or traits, homework, and extra credit. Also similar to elementary teachers, a clear component for middle and high school teachers grouped items that involved constructed response assessments, such as essays, performance-based, and projects. Other components for type of assessment used were object tests and quizzes, how assessments are constructed (by publisher or the teacher), and for middle school teachers, major exams. Finally, with respect to cognitive levels of assessments, understanding, reasoning and higher order items formed one factor, with recall knowledge a second factor for high school teachers. Overall, then, there were very few differences between middle and high school teachers, as well as considerable similarity to what was found for elementary teachers.

Relationship Results To examine the relationship of assessment and grading practices to grade level, subject matter, and ability level of students, several analyses were completed using component scores as dependent variables. The results of these ANOVAs are summarized in

Table 17 for both middle and high school teachers. Table 18 shows the means of selected items from each component that resulted in significant differences.

While there were many statistically significant differences found in the analyses, it is noteworthy to point out that there were few clear trends or relationships that are inconsistent with common practice. Overall, most of the means broken out by grade level, subject, and ability level were not different. Only half of the total number of ANOVAs computed (10) showed any statistical significance, and post hoc tests indicated that, with two exceptions, the differences were confined to one ability level or subject compared to one other. This points out that the vast majority of paired comparisons did not show significance. Only one component, academic performance, showed any significance by grade level. The results show that in middle school there is a positive relationship between grade level and the emphasis placed on academic performance. However grade level and academic performance were not related in high school.

With respect to differences according to ability level of the classes, as might be expected, teachers of advanced classes emphasize academic performance, constructed response assessments, major exams, and reasoning more than standard or basic classes, while basic classes emphasize homework and extra credit more than advanced classes.

In examining differences between subjects at the middle school level, there were only two components that showed significance for middle school. Math classes emphasized enabling-performance behaviors such as effort and improvement more than English classes. Math classes also utilized constructed response items less than English, social studies, or science classes. At the high school level there were more significant differences. Social studies classes emphasized academic performance less than math classes, and emphasized extra credit more than English classes. With respect to factors, academic performance is emphasized more in math than in

social science while extra credit is emphasized more in social studies than in English. With types of assessments, English teachers used more constructed response assessments than the other three subjects, and social studies more than math. English teachers used assessments prepared by themselves more than science or math teachers, and English teachers used assessments of reasoning skills more than the other three subjects. English teachers also used recall items less than math or science, social studies less than math, and science less than social studies.

Table 17

Relationship of component scores with ability level of class, subject matter and grade level

New Factors by Ability Level	p value	Relationship
Middle School		
Factor 2 between advanced/honors and standard	.007	AH>Std
Factor 2 between advanced/honors and basic	.003	AH>Basic
Type 1 between advanced/honors and standard	.001	AH>Std
Type 4 between advanced/honors and standard	.000	AH>Std
Type 4 between advanced/honors and basic	.000	AH>Basic
CA 1 between advanced/honors and basic	.000	AH>Basic
CA 1 between advanced/honors and standard	.000	AH>Std
High School		
Factor 2 between advanced/honors and standard	.013	AH>Std
Factor 2 between advanced/honors and basic	.002	AH>Basic
Factor 3 between advanced placement and Standard	.019	Std>AP
Factor 3 between advanced placement and basic	.008	Basic>AP
Factor 4 between advanced/honors and basic	.041	Basic>AH
Type 1 between advanced/honors and basic	.006	AH>Basic
Type 1 between advanced/honors and standard	.003	AH>Basic
Type 2 between advanced/honors and basic	.000	AH>Basic
Type 2 between advanced/honors and standard	.024	AH>Std

Type 2 between basic and standard	.014	Std>Basic
CA 1 between advanced placement and standard	.000	AP>Std
CA1 between advanced placement and basic	.000	AP>Basic
CA 1 between advanced/honors and standard	.001	AH>Std
CA 1 between advanced/honors and basic	.000	AH>Basic
CA 1 between standard and basic	.005	Std>basic

New Factors by Subject Matter

Middle School

Factor 1 between math and English	.035	Math>Eng
Type 1 between math and English	.000	Eng>Math
Type 1 between math and science	.000	Sci>Math
Type 1 between math and history/ss	.000	SS>Math

High School

Factor 2 between math and history/ss	.021	Math>SS
Factor 4 between English and history/ss	.010	SS>Eng
Type 1 between English and science	.000	Eng>Sci
Type 1 between English and history/ss	.002	Eng>SS
Type 1 between English and math	.000	Eng>math
Type 1 between math and science	.000	Sci>Math
Type 1 between math and history/ss	.000	SS>Math
Type 2 between English and science	.001	Eng>Science

Type 2 between English and math	.018	Eng>Math
CA 1 between English and science	.006	Eng>Sci
CA 1 between English and math	.000	Eng>Math
CA 1 between English and history/ss	.001	Eng>SS
CA 1 between science and history/ss	.019	Sci>SS
CA2 between English and history/ss	.000	SS>Eng*
CA2 between English and math	.033	Math>Eng
CA2 between English and science	.047	Sci>Eng
CA2 between math and history/ss	.004	SS>Math
CA2 between science and history/ss	.002	Sci>SS

New Factors by Grade Level

Middle School

Type 2 between grades 6 and 8	.036	6>8
Type 2 between grades 7 and 8	.001	7>8

High School

No significant differences found

Key: AP=advanced placement; AH=advanced/honors; Basic=basic; Std=standard; Eng=English; Math=math; Sci=science; SS=social studies/history.

The relationship between the components and grades awarded was examined by regression analyses, using percent As awarded as the dependent variable and the component scores as independent variables. The results of these analyses are summarized in Table 19. These results indicate that few of the components are related to how many As are awarded. For middle school, a positive relationship exists between the emphasis placed on constructed response assessments and As awarded. In high school, three components predicted percentage of As awarded. The use of zeros was related to As awarded.

Table 18

Components as Predictors of Percentage As Awarded by Middle and High School Teachers

New Components	R	P value
Middle School	.185	
Constructed response assessments		.010
High School	.265	
Use of 0s in calculating grades		.006
Objective assessments		.006
Use of recall items		.000

Summary of Secondary Level Findings

The results of the secondary level analyses are summarized as follows:

- Like elementary teachers, most secondary teachers use a multitude of factors in grading students.
- Few differences exist between middle school and high school teachers concerning factors used in grading, types of assessments used, or cognitive levels of assessments. Middle school teachers do tend to weight effort more than high school teachers.
- Disruptive student behavior, grade distributions of other teachers, performance compared to students from previous years, school division policy about the percentage of students who may obtain different grades, and extra credit for non academic performance contribute little to determining grades.
- Academic performance, performance compared to a set scale of percentage correct, and specific learning objectives mastered contribute most to the determination of grades.
- Student effort, ability levels of students, quality of homework completed, class participation and attendance, and inclusion of zeros in calculating grades contribute moderately to final semester grades.
- Relatively large variation in the emphasis of moderately important factors suggests that some teachers weight these factors significantly while other teachers place much less emphasis on them.
- About half the secondary teachers indicated that they use performance compared to a set scale of percentage correct either extensively or completely.

- More emphasis is placed on assessments of understanding, reasoning, and higher order thinking skills than recall, but a substantial percentage of teachers still use recall items quite a bit or extensively.
- Secondary teachers use of wide variety of assessments. Objective assessments are used only slightly more than constructed response type assessments such as essays, performance-based assessments, and projects.
- A significant percentage of teachers indicate a need for in-service in several areas, especially using assessment during instruction and improving the overall quality of classroom tests, understanding SOL tests, assessing reasoning, and understanding the link between assessment and instruction.
- About half the grades given in middle school and grade 12 are As and Bs; 36% in grades 10 and 11.
- Secondary teachers emphasize four major factors in grading: academic performance, enabling performance behaviors, homework, and extra credit. Most emphasis is placed on the first two components.
- Assessments that require constructed responses, such as performance-based assessments, comprised a component among items addressing types of assessments used.
- Reasoning, understanding, and higher order thinking tend to be clustered together, and, for high school teachers, considered separate from items measuring recall.
- Only the emphasis on academic performance was related to grade level.

- Advanced and AP classes showed significantly more emphasis, as compared to standard and basic classes, on academic performance, use of constructed response assessments, use of major exams, items assessing reasoning, and assessments designed by themselves, and emphasized zeros and extra credit less.
- Math teachers emphasized constructed response and recall items less than English, social science and science teachers, and academic performance more than social science teachers. English teachers emphasized reasoning and teacher-made assessments more. Otherwise, few differences based on grade level, subject, or ability level of the class were revealed.
- There were few relationships between assessment and grading practices and grades awarded. Teachers who award more As tend to be those at the middle school who use constructed response assessments more and those at the high school that use zeros in calculating grades less and objective, recall test items more.

Conclusions

The results of these analyses of a large sample of teachers indicates that teachers use a variety of factors assessing and grading students, with different teachers weighting these factors in idiosyncratic ways. Two factors appear to have the greatest influence on determining grades, academic performance and achievement, and behaviors and traits that are related to performance, such as effort, ability, and participation, are important contributors to determining grades. Two other factors, homework and extra credit, are less important. A consistent finding in examining the assessment and grading practices is that there is great variation among teachers in how much different practices are used and the contribution of different factors to determine grades. Little of

the variation can be explained by grade level, subject matter, or ability level of the class, suggesting that teachers may develop idiosyncratic practices based to only a small extent on grade level, subject or ability levels of the students. Most of the variation in practice occurs with factors that have a moderate influence on grades, such as effort, participation, homework, and improvement.

Teachers at all levels indicated significant needs for professional development in several assessment areas, including how to use assessment during instruction, improving the quality of classroom tests, understanding SOL tests and using SOL test results, assessing reasoning, and, for elementary teachers, assessing reading and writing.

At the high school level it appears that advanced and AP classes use more constructed response assessments and tend to focus more on reasoning and other higher order thinking skills. These teachers also emphasized academic performance more, assessments designed by themselves. Extra credit and zeros were used less. While high school teachers indicate they use objective assessment most overall, constructed response assessments, such as performance-based assessments, essays, and projects, are used extensively. Math teachers tend to emphasize constructed response items less than teachers of other areas, while English teachers emphasize reasoning and teacher made assessments more.

Only a small relationship was found between assessment and grading practices and grades awarded. There was a tendency for elementary teachers who used more assessments of reasoning and less objective items and homework, for middle school teachers who used constructed response items more, and high school teachers who assess recall knowledge to award more As.

Implications

The finding from this study suggest several implications for teachers, staff development, and administrators.

1. The wide variation of assessment and grading practices within each school suggests a need to examine if stated grading policies and procedures are accurate. Considering the significant contribution for many teachers of factors such as effort, improvement, and participation, is it clear how these factors are incorporated? How do teachers monitor and "grade" effort? Also, is it acceptable, or desirable, to maintain the essentially private, idiosyncratic approach to assessment and grading that results in such wide variation? Would it be helpful to have discussions among teachers concerning the weight given to the performance-enabling factors and how they are documented?
2. Currently, constructed response assessment and assessment of reasoning and higher order thinking skills is used extensively. With the emphasis on SOL tests it may be both interesting and informative to monitor the extent to which these kind of assessments are continued.
3. Teachers have indicated a need for further professional development on many assessment issues and techniques, which suggests that efforts to provide such training would be welcome. This topic is one that can be addressed with new SOL Training Initiative funds. At the very least, it would seem important to make sure teachers have access to information and materials to help them improve their assessment and grading skills. The results of the survey identify those areas that teachers view as most important for professional development.

4. A significant percentage of teachers use zeros in calculating grades. This is a sometimes contentious issue on which there is little consensus. It may be helpful to identify alternatives to including zeros in calculating grades and to explore whether district or school policies concerning this practice should be developed.
5. To what extent do teachers differentiate understanding from reasoning and other higher order skills? The data from this study suggest that understanding may not be differentiated, while the literature clearly does emphasize understanding as different from reasoning. It would be helpful to explore this issue further with teachers, as well as to examine assessments they give to students to determine what cognitive levels are being tested. It is also interesting that there is a high emphasis on assessing recall knowledge. Further exploration of what is meant by recall knowledge may be helpful in bringing attention to the difference between knowing and understanding.
6. Given that assessment and grading practices are not well described by grade level, subject, or ability level of the students, what does influence these practices? This question can be investigated further through interviews of teachers and through cluster analysis statistical procedures of the current data.

References

Airasian, P. W. (1997). Classroom assessment (3rd Edition). NY: McGraw-Hill.

Airasian, P. W. (1984). Classroom assessment and educational improvement. Paper presented at the conference Classroom Assessment: A Key to Educational Excellence, Northwest Regional Educational Laboratory, Portland, Oregon.

American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, NEA). (1990) Standards for teacher competence in educational measurement. Washington, DC:. Author.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology, 84, 261-271.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. Applied Measurement in Education, 10, 161-180.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. Applied Measurement in Education, 7, 279-301.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. Journal of Educational Measurement, 30, 123-142.

Brookhart, S. M. (1991). Grading practices and validity. Educational Measurement: Issues and Practice, 10, 35-36.

Carter, K. (1984). Do teachers understand the principles for writing tests? Journal of Teacher Education, 35, 57-60.

Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye, Editor, Handbook of classroom assessment, NY: Academic Press.

Cizek, G. J., Fitzgerald, Shawn M., & Rachor, Robert E. (1995). Teachers' Assessment Practices: Preparation, Isolation and the Kitchen Sink. Educational Assessment, 3(2), 159-179.

Cross, L. H., & Frary, R. B. (1996). Hodgepodge grading: Endorsed by students and teachers alike. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway, ed., Testing in the schools. New directions for testing and measurement. San Francisco: Jossey-Bass.

Frary, R.B., Cross, L.H. & Weber, L.J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice 12(3), 23-30.

Friedman, S. J. & Manley, M. (1991). Grading practices in the secondary school: Perceptions of the stakeholders. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago.

Haertel, E., Ferrara, S., Korpi, M., & Prescott, B. (1984). Testing in secondary schools: Student perspectives. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

McMillan, J. H. (1997). Classroom assessment: Principles and practice for effective instruction. Boston: Allyn & Bacon.

Messick, S. (1989). Validity. In R. L. Linn (ed.), Educational measurement (3rd Edition). Washington, DC: American Council on Education and National Council on Measurement in Education.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62, 229-258.

Phye, G. D. (1997). Classroom assessment: A multidimensional perspective. In G. D. Phye, Editor, Handbook of classroom assessment, NY: Academic Press.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye, ed., Handbook of classroom assessment, NY: Academic Press.

Plake, B. S., & Impara, J. C. (1993). Assessment competencies of teachers: A national survey, Educational Measurement: Issues and Practice, 12, 10-25.

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Shunk & J. L. Meece (Eds.), Student perceptions in the classroom (pp. 149-183). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pintrich, P. R., & Schunk, D. H. (1996). Motivation in education: Theory, research, and applications, Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J. (1995). Classroom assessment: What teachers need to know, Boston: Allyn & Bacon.

Schafer, W. D. (1993). Assessment in teacher education. Theory into Practice, 32, 118-126.

APPENDICES

Survey of Assessment and Grading Practices - Elementary Form

Directions: Answer the questions by circling the number that most closely corresponds to your assessment and grading practices for your class **first semester** in mathematics and language arts/reading. There are no right or wrong answers; all your responses will be kept confidential.

For questions 1- 34 use the following response scale:

1	2	3	4	5	6
Not at All	*Very Little*	*Some*	*Quite a Bit*	*Extensively*	*Completely*

To what extent were final first semester grades of students in your class based on:	For Mathematics						For Language Arts/Reading					
1. disruptive student behavior	1	2	3	4	5	6	1	2	3	4	5	6
2. improvement of performance since the beginning of the year	1	2	3	4	5	6	1	2	3	4	5	6
3. student effort - how much the student tried to learn	1	2	3	4	5	6	1	2	3	4	5	6
4. ability levels of the students	1	2	3	4	5	6	1	2	3	4	5	6
5. major exams	1	2	3	4	5	6	1	2	3	4	5	6
6. oral presentations	1	2	3	4	5	6	1	2	3	4	5	6
7. work habits and neatness	1	2	3	4	5	6	1	2	3	4	5	6
8. grade distributions of other teachers	1	2	3	4	5	6	1	2	3	4	5	6
9. completion of homework (not graded)	1	2	3	4	5	6	1	2	3	4	5	6
10. the quality of completed homework	1	2	3	4	5	6	1	2	3	4	5	6
11. academic performance as opposed to other factors	1	2	3	4	5	6	1	2	3	4	5	6
12. performance compared to other students in the class	1	2	3	4	5	6	1	2	3	4	5	6
13. performance compared to a **set scale of percentage correct** (e.g., 86-94% B)	1	2	3	4	5	6	1	2	3	4	5	6
14. performance compared to students from **previous years**	1	2	3	4	5	6	1	2	3	4	5	6
15. specific learning objectives mastered	1	2	3	4	5	6	1	2	3	4	5	6
16. formal or informal school or district policy of the percentage of students who may obtain As, Bs, Cs, Ds, and Fs.	1	2	3	4	5	6	1	2	3	4	5	6

GO ON TO NEXT PAGE ⟶

1 *Not At All* 2 *Very Little* 3 *Some* 4 *Quite a Bit* 5 *Extensively* 6 *Completely*

To what extent were final first semester grades of students in your class based on:	For Mathematics	For Language Arts/Reading
17. the degree to which the student pays attention and/or participates in class	1 2 3 4 5 6	1 2 3 4 5 6
18. the inclusion of 0s for incomplete assignments in the determination of final percentage correct	1 2 3 4 5 6	1 2 3 4 5 6
19. assessments that measure student recall knowledge	1 2 3 4 5 6	1 2 3 4 5 6
20. assessments that measure student understanding	1 2 3 4 5 6	1 2 3 4 5 6
21. assessments that measure student reasoning (higher order thinking)	1 2 3 4 5 6	1 2 3 4 5 6
22. assessments that measure how well students apply what they learn	1 2 3 4 5 6	1 2 3 4 5 6
23. objective assessments (e.g., multiple choice, matching, short answer)	1 2 3 4 5 6	1 2 3 4 5 6
24. performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	1 2 3 4 5 6	1 2 3 4 5 6
25. assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	1 2 3 4 5 6	1 2 3 4 5 6
26. assessments designed primarily by yourself	1 2 3 4 5 6	1 2 3 4 5 6
27. essay-type questions	1 2 3 4 5 6	1 2 3 4 5 6
28. projects completed by teams of students	1 2 3 4 5 6	1 2 3 4 5 6
29. projects completed by individual students	1 2 3 4 5 6	1 2 3 4 5 6
30. performance on quizzes	1 2 3 4 5 6	1 2 3 4 5 6
31. extra credit for nonacademic performance (e.g., bringing in items for food drive)	1 2 3 4 5 6	1 2 3 4 5 6
32. extra credit for academic performance	1 2 3 4 5 6	1 2 3 4 5 6
33. effort, improvement, behavior, and other "nontest" indicators **for borderline cases**	1 2 3 4 5 6	1 2 3 4 5 6
34. authentic assessments (i.e., "real world" performance tasks)	1 2 3 4 5 6	1 2 3 4 5 6

GO ON TO NEXT PAGE ⟶

35. What is the ***approximate*** percentage of different letter grades given to your class for the first semester in <u>mathematics</u> and <u>language arts</u>? (i.e., what percent of the grades given were As, Bs, etc.)

Mathematics		Language Arts	
A (A+, A, A-)	_____ %	**A** (A+, A, A-)	_____ %
B (B+, B, B-)	_____ %	**B** (B+, B, B-)	_____ %
C (C+, C, C-)	_____ %	**C** (C+, C, C-)	_____ %
D (D+, D, D-)	_____ %	**D** (D+, D, D-)	_____ %
F	_____ %	**F**	_____ %
	100% (Total must add to 100)		**100%** (Total must add to 100)

Check here if you did not use letter grades.

Questions 36 - 56. Indicate the ***importance*** of each of the following potential **STAFF DEVELOPMENT** topics **for you** by circling the appropriate response:

Use the following response scale:

1	2	3	4	5
Not at All Important	***Of Little Importance***	***Somewhat Important***	***Very Important***	***Critical***

36. Using assessment information **for planning prior to instruction**	1	2	3	4	5
37. Using assessment information **during instruction** (e.g., monitoring student progress, judging whether students understand, questioning students)	1	2	3	4	5
38. Using assessment results to evaluate instruction and curriculum	1	2	3	4	5
39. Using assessment results to determine student grades	1	2	3	4	5
40. Communicating with parents concerning grades and test scores	1	2	3	4	5
41. Understanding and using the new Stanford 9 standardized tests	1	2	3	4	5
42. Understanding and using the new SOL tests	1	2	3	4	5
43. Understanding technical assessment concepts such as reliability and validity	1	2	3	4	5
44. Improving the overall quality of classroom assessments	1	2	3	4	5
45. Assessing reasoning and other "higher order" thinking skills	1	2	3	4	5
46. Using performance-based assessments, such as presentations and projects	1	2	3	4	5

	1 *Not At All Important*	2 *Of Little Importance*	3 *Somewhat Important*	4 *Very Important*	5 *Critical*
47. Using portfolio assessments	1	2	3	4	5
48. Designing paper and pencil tests (e.g., multiple choice, short answer, essay)	1	2	3	4	5
49. Assessing writing skills	1	2	3	4	5
50. Assessing reading proficiency	1	2	3	4	5
51. Assessing mainstreamed students	1	2	3	4	5
52. Assessing affective traits, such as attitudes, values, and self-concept	1	2	3	4	5
53. Understanding the link between assessment and instruction	1	2	3	4	5
54. Calculating final course or semester grades	1	2	3	4	5
55. Other: _____	1	2	3	4	5
56. Other: _____	1	2	3	4	5

57. Have the new Virginia Standards of Learning (SOLs) impacted on your **assessment** or **grading** of students? (Circle one)

Yes, extensively **Yes, somewhat** **Yes, very little** **No**

If Yes, briefly describe the change(s): _____

_____

_____

_____

_____

58. Grade Level of Class (Circle one) **3** **4** **5** **3-4** **4-5** **3-5**

59. Name of School: _____

60. Your Name: _____
(**Optional** - names will be used **only** for selected follow-up interviews)

THANK YOU FOR YOUR COOPERATION!

Survey of Assessment and Grading Practices - Secondary Form

Directions: Answer the questions by keeping in mind your assessment and grading practices for a **single class** you taught first semester. The class should be the most typical section of the course you teach most frequently. **Answer all the questions with this class in mind**. There are no right or wrong answers; all your responses will be kept confidential.

1. Grade level of class (check one): ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12
   ☐ combination of two or more grades

2. Subject (check one): ☐ mathematics ☐ English ☐ science ☐ history/social science ☐ other

3. Ability level of class (check one): ☐ AP ☐ advanced/honors ☐ standard ☐ basic ☐ mixed ☐ other

For questions 4 - 37 use the following response scale to circle your answers:

1	2	3	4	5	6
Not at All	*Very Little*	*Some*	*Quite a Bit*	*Extensively*	*Completely*

To what extent were final first semester grades of students <u>in your single class described above</u> based on:

4. disruptive student behavior	1	2	3	4	5	6
5. improvement of performance since the beginning of the year	1	2	3	4	5	6
6. student effort - how much the student <u>tried</u> to learn	1	2	3	4	5	6
7. ability levels of the students	1	2	3	4	5	6
8. major exams	1	2	3	4	5	6
9. oral presentations	1	2	3	4	5	6
10. work habits and neatness	1	2	3	4	5	6
11. grade distributions of other teachers	1	2	3	4	5	6
12. completion of homework (not graded)	1	2	3	4	5	6
13. the quality of completed homework (graded)	1	2	3	4	5	6
14. academic performance as opposed to other factors	1	2	3	4	5	6
15. performance compared to other students in the class	1	2	3	4	5	6
16. performance compared to **a set scale of percentage correct** (e.g., 86-94% B)	1	2	3	4	5	6

GO ON TO NEXT PAGE ⟶

1	2	3	4	5	6
Not At All	*Very Little*	*Some*	*Quite a Bit*	*Extensively*	*Completely*

To what extent were final first semester grades of students <u>in your single class described above</u> based on:

17.	performance compared to students from **previous years**	1	2	3	4	5	6
18.	specific learning objectives mastered	1	2	3	4	5	6
19.	formal or informal school or district policy of the percentage of students who may obtain As, Bs, Cs, Ds, and Fs.	1	2	3	4	5	6
20.	the degree to which the student pays attention and/or participates in class	1	2	3	4	5	6
21.	the inclusion of 0s for incomplete assignments in the determination of final percentage correct	1	2	3	4	5	6
22.	assessments that measure student <u>recall</u> knowledge	1	2	3	4	5	6
23.	assessments that measure student understanding	1	2	3	4	5	6
24.	assessments that measure student reasoning (higher order thinking)	1	2	3	4	5	6
25.	assessments that measure how well students apply what they learn	1	2	3	4	5	6
26.	objective assessments (e.g., multiple choice, matching, short answer)	1	2	3	4	5	6
27.	performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	1	2	3	4	5	6
28.	assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	1	2	3	4	5	6
29.	assessments designed primarily by yourself	1	2	3	4	5	6
30.	essay-type questions	1	2	3	4	5	6
31.	projects completed by teams of students	1	2	3	4	5	6
32.	projects completed by individual students	1	2	3	4	5	6
33.	performance on quizzes	1	2	3	4	5	6
34.	extra credit for nonacademic performance (e.g., bringing in items for food drive)	1	2	3	4	5	6
35.	extra credit for academic performance	1	2	3	4	5	6
36.	effort, improvement, behavior, and other "nontest" indicators for **borderline** cases	1	2	3	4	5	6
37.	authentic assessments (i.e., "real world" performance tasks)	1	2	3	4	5	6

GO ON TO NEXT PAGE ⟶

38. What is the ***approximate*** percentage of different letter grades given **in the class you selected above** for the first semester? (i.e., what percent of the grades given were As, Bs, etc. for this class?)

A (A+, A, A-)	_____ %
B (B+, B, B-)	_____ %
C (C+, C, C-)	_____ %
D (D+, D, D-)	_____ %
F	_____ %
	100% (Total must add to 100)

Questions 39 - 59. Indicate the ***importance*** of each of the following potential **STAFF DEVELOPMENT** topics **for you** by circling the appropriate response.

Use the following response scale:

1	2	3	4	5
Not at All Important	***Of Little Importance***	***Somewhat Important***	***Very Important***	***Critical***

39. Using assessment information **for planning prior to instruction**	1	2	3	4	5
40. Using assessment information **during instruction** (e.g., monitoring student progress, judging whether students understand, questioning students)	1	2	3	4	5
41. Using assessment results to evaluate instruction and curriculum	1	2	3	4	5
42. Using assessment results to determine student grades	1	2	3	4	5
43. Communicating with parents concerning grades and test scores	1	2	3	4	5
44. Understanding and using the new Stanford 9 standardized tests	1	2	3	4	5
45. Understanding and using the new SOL tests	1	2	3	4	5
46. Understanding technical assessment concepts such as reliability and validity	1	2	3	4	5
47. Improving the overall quality of classroom assessments	1	2	3	4	5
48. Assessing reasoning and other "higher order" thinking skills	1	2	3	4	5
49. Using performance-based assessments, such as presentations and projects	1	2	3	4	5
50. Using portfolio assessments	1	2	3	4	5
51. Designing paper and pencil tests (e.g., multiple choice, short answer, essay)	1	2	3	4	5
52. Assessing writing skills	1	2	3	4	5

GO ON TO NEXT PAGE ⟶

	1 *Not At All Important*	2 *Of Little Importance*	3 *Somewhat Important*	4 *Very Important*	5 *Critical*
53. Assessing reading proficiency	1	2	3	4	5
54. Assessing mainstreamed students	1	2	3	4	5
55. Assessing affective traits, such as attitudes, values, and self-concept	1	2	3	4	5
56. Understanding the link between assessment and instruction	1	2	3	4	5
57. Calculating final course or semester grades	1	2	3	4	5
58. Other: _____	1	2	3	4	5
59. Other: _____	1	2	3	4	5

60. Have the new Virginia Standards of Learning (SOLs) impacted on your **assessment** or **grading** of students? (Circle one)

Yes, extensively **Yes, somewhat** **Yes, very little** **No**

If Yes, briefly describe the change(s): _____

_____

_____

_____

_____

61. Name of School: _____

62. Your Name: _____
(**Optional** - names will be used **only** for selected follow-up interviews)

THANK YOU FOR YOUR COOPERATION!

TEACHERS' CLASSROOM ASSESSMENT AND GRADING PRACTICES:

Phase 2

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

CHESTERFIELD COUNTY PUBLIC SCHOOLS • COLONIAL HEIGHTS CITY SCHOOLS • HANOVER COUNTY PUBLIC SCHOOLS • HENRICO COUNTY PUBLIC SCHOOLS • HOPEWELL CITY PUBLIC SCHOOLS • POWHATAN COUNTY PUBLIC SCHOOLS • RICHMOND CITY PUBLIC SCHOOLS • VIRGINIA COMMONWEALTH UNIVERSITY

METROPOLITAN EDUCATIONAL RESEARCH CONSORTIUM

PO BOX 842020 • RICHMOND VA 23284-2020 • Phone: (804) 828-0478
FAX: (804) 828-0479 • E-MAIL: jmcmilla@saturn.vcu.edu

MERC MEMBERSHIP
James H. McMillan, Director

CHESTERFIELD COUNTY PUBLIC SCHOOLS
William C. Bosher, Jr., Superintendent

COLONIAL HEIGHTS CITY SCHOOLS
James L. Ruffa, Superintendent

HANOVER COUNTY PUBLIC SCHOOLS
Steward D. Roberson, Superintendent

HENRICO COUNTY PUBLIC SCHOOLS
Mark A. Edwards, Superintendent

HOPEWELL CITY PUBLIC SCHOOLS
David C. Stuckwisch, Superintendent

POWHATAN COUNTY PUBLIC SCHOOLS
Margaret S. Meara, Superintendent

RICHMOND CITY PUBLIC SCHOOLS
Albert J. Williams, Superintendent

VIRGINIA COMMONWEALTH UNIVERSITY
John S. Oehler, Dean
School of Education

Virginia Commonwealth University and the school divisions of Chesterfield, Colonial Heights, Hanover, Henrico, Hopewell and Richmond established the **Metropolitan Educational Research Consortium** (MERC) on August 29, 1991. The founding members created **MERC** to provide timely information to help resolve educational problems identified by practicing professional educators. **MERC** membership is open to all metropolitan-type school divisions. It currently provides services to 9,000 teachers and 138,000 students. **MERC** has base funding from its membership. Its study teams are composed of University investigators and practitioners from the membership.

MERC is organized to serve the interests of its members by providing tangible material support to enhance the practice of educational leadership and the improvement of teaching and learning in metropolitan educational settings. **MERC's** research and development agenda is built around four goals:

- To improve educational decision-making through joint development of practice-driven research questions, design and dissemination,
- To anticipate important educational issues and provide leadership in school improvement,
- To identify proven strategies for resolving instruction, management, policy and planning issues facing public education, and
- To enhance the dissemination of effective school practices.

In addition to conducting research as described above, **MERC** will conduct technical and issue seminars and publish reports and briefs on a variety of educational issues.

TEACHERS' CLASSROOM ASSESSMENT AND GRADING PRACTICES:

Phase 2

James H. McMillan, Professor
Virginia Commonwealth University

Daryl Workman, MERC Research Fellow
Virginia Commonwealth University

Copyright©1999. Metropolitan Educational Research Consortium (MERC), Virginia Commonwealth University

*The views expressed in MERC publications are those of individual authors and not necessarily those of the Consortium or its members.

Executive Summary

Classroom assessment and grading practices are becoming a greater focus of educational inquiry as teachers and policymakers become more accountable to the public for educational outcomes via assessment driven instructional practices. This study is an attempt to better understand the classroom assessment and grading practices of teachers, which have previously been described as a hodgepodge mix of student attitude, effort and achievement. Specifically, the following questions regarding teachers' assessment and grading practices are addressed:

- What is the current state of assessment practice and grading by teachers?
- What assessment and grading topics do teachers identify as needs to be addressed in in-service?
- What is the relationship between assessment and grading practices and grades given to students?
- What are the relationships between grade level, and subject taught and assessment and grading practices?
- What are the reasons teachers give for their assessment and grading decision-making?
- What is the impact of the SOL tests on the extent to which different assessment techniques are used in the classroom?
- What classroom assessment and in-service needs do teachers have?

Results of the research indicates that teachers do in fact use a multitude of factors to assess and grade students, including academic performance, effort, improvement, ability, homework, and extra credit. However, this study attempts to look beyond a "hodgepodge" explanation of assessment and grading practices in order to uncover relationships that help to further explain teachers' assessment and grading practices and decision-making processes.

This report summarizes the findings from Phase 2 of the study, which focused on personal interviews with teachers. Phase 1 of this research, which was reported in Fall, 1998, surveyed 921 elementary, 597 middle and 850 high school teachers. A total of 28 mostly middle and high school mathematics and English teachers were interviewed individually.

The analysis of Phase 2 data indicates that there is tension between two sources of influence on teacher decision-making concerning assessment and grading practices. One source is teacher beliefs and values while a second source is external pressures and constraints. Teacher beliefs and values focus on assessment and grading practices that will encourage and support student learning. Teachers "pull" for students, devising approaches to assessment and grading that make it likely that students will succeed. Assessment and grading practices tend to be individualized to a certain extent for different students, and used as a way to keep students motivated and engaged. Teachers want students to understand and learn, and want assessments that help this outcome. Constructed-response assessments are seen as providing the best information to help students succeed.

Outside pressures and constraints include parental demands and informing parents of student progress, division policies, skills needed by students once they graduate, practical constraints such as having over 100 students, and perhaps most importantly, state mandated high stakes multiple choice testing. It appears that the state testing program has become a significant influence on teacher decision making, lessening to some extent assessment and grading practices that more clearly, from the teachers' perspectives, promote student learning.

Implications for teacher professional training and development are made in light of the tension between these two sources influences.

Table of Contents

Preface 1

Introduction 2

Review of Literature 4

Research Questions 13

Methodology 14

Findings 20

Conclusions 44

Implications 46

References 49

Appendix A 52

Preface

The research in this report was directed by a team of individuals. This team identified the research problem and questions, developed a research design, assisted in gathering data from teachers, and took an active role in identifying samples and analyzing data. The principal investigators are grateful for their contribution and assistance. The members of the research team include the following:

James McMillan
Virginia Commonwealth University

Daryl Workman
Virginia Commonwealth University

Lin Corbin-Howerton
Chesterfield County Public Schools

Joseph Tylus
Monacan High School
Chesterfield County Public Schools

Ann Williams
Colonial Heights Middle School
Colonial Heights Public Schools

James Bagby
Hanover County Public Schools

Carole Urbansok-Eads
Hanover County Public Schools

Deborah Pittman
Cold Harbour Elementary School
Hanover County Public Schools

Catherine Nolte
Henrico County Public Schools

Yvonne Smith-Jones
Hopewell City Public School

Stephanie Couch
Pocahontas Middle School
Powhatan County Public Schools

Charmaine Brooks
Westover Hills Elementary
Richmond City Public Schools

Pat Janes
Armstrong High School
Richmond City Public Schools

Sue Jones
Patrick Copeland Elementary
Richmond City Public Schools

Audrey Johnson
Thomas Jefferson High School
Richmond City Public Schools

Richard Williams
Richmond City Public Schools

Introduction

A significant amount of recent literature has focused on classroom assessment and grading as essential aspects of effective teaching. There is an increased scrutiny of assessment as indicated by the popularity of performance assessment and portfolios, newly established national assessment competencies for teachers (Standards, 1990), and the interplay between learning, motivation, and assessment (Brookhart, 1993, 1994; Tittle, 1994). In Virginia, the Standards of Learning and associated tests highlight the importance of assessment.

Previous research documents that teachers tend to award a "hodgepodge grade of attitude, effort, and achievement" (Brookhart, 1991, p. 36). It is also clear that teachers use a variety of assessment techniques, even if established measurement principles are often violated (Cross & Frary, 1996; Frary, Cross, & Weber, 1993; Plake & Impara, 1993; and Stiggins & Conklin, 1992).

Given the variety of assessment and grading practices in the field, the increasing importance of assessment, the critical role each classroom teacher plays in determining assessments and grades, and the trend toward greater accountability of teachers with state assessment approaches that are inconsistent with much of the current literature, there is a need to (1) understand current assessment and grading practices, (2) understand the relationship of these practices to grades given by teachers, (3) determine if "standards" teachers use to assign grades differ from one classroom to another and one school to another, (4) examine the consequential validity of the

new SOL tests on classroom assessment practices, and (5) determine assessment and grading topics that, according to teachers, need in-service.

The fourth need is related to a recently expanded conception of test validity that includes what has been called "consequential validity" or "consequential bias" (Messick, 1989; Moss, 1992). Essentially, test developers and users need to be sensitive to how assessments influence instructional practices and curriculum. The importance of consequential validity is indicated by its inclusion in the new Standards for Educational and Psychological Testing. Of interest in the current study is the effect the new statewide assessment program may have on instructional practices. For example, the assessments may result in teachers stressing a particular method of instruction or classroom testing that is consistent with the emphasis and approach adopted in the statewide system

There is a need to provide information that addresses issues of consistency and fairness in assessment and grading across classrooms and schools, to illustrate to teachers the nature of current practice and provide a stimulus for discussion, and to establish assessment and grading policy. There is also a need to understand the motivation and reasons for using specific assessment and grading practices.

The purpose of both Phases of this investigation, then, was to describe the classroom assessment and grading practices of teachers, determine if meaningful relationships exist between these practices and grade level, subject matter, ability levels of different classes, and to understand the reasons teachers give for using certain assessment and grading practices, and to document teacher needs for inservice education related to assessment. This report focuses on Phase 2 of the research.

Review of Literature

Despite the growing importance of classroom assessment and the introduction of new methods of assessment, there is relatively little research on the nature and effects of classroom assessments on student learning and motivation (Stiggins, 1997). Most assessment research has focused on standardized testing, despite evidence that teachers spend considerable time assessing students, and that student well-being is influenced by the quality of assessments given by the teacher (Stiggins and Conklin, 1992). Also, there is little empirical research on classroom assessments, with measurement experts tending instead to pay much more attention to large scale testing than classroom assessment. It is also evident that many teachers lack assessment competency (Plake and Impara, 1997). This isn't too surprising, however, since less than 50% of the teacher certification programs in the United States require no measurement course (Schafer, 1993). This remains the case, despite the fact that teacher standards for assessment competency were identified in 1990 (AFT, NCME, NEA, 1990).

Prior to the mid 1980s the literature on educational assessment focused almost exclusively on large-scale standardized testing. According to Stiggins and Conklin (1992), most inquiry on classroom assessment was based on a conceptualization similar to what had been developed for standardized testing, emphasizing paper and pencil, multiple choice testing. Furthermore, the only written standards for assessment, *Standards for Educational and Psychological Testing*, dealt primarily with standardized tests. Finally, during the 1980s the emerging literature about teacher decision-making, teacher behavior, and student achievement found little on how

classroom assessments relate to teaching or learning. Shulman (1980) concluded that most of the paper and pencil tests used for assessment were inconsistent with, and often irrelevant to, the realities of teaching. Haertel, et al. (1984), in a review of research on high school testing, concluded that little is known about teachers' or students' perceptions of the impacts of classroom assessment.

Phye (1997) states that "it is not only the assessment option that determines what we get as evidence of learning or achievement. How we use the assessment instruments or techniques also determine the nature of the knowledge a student is demonstrating. *How* we assess determines *what* we get" and thus classroom learning and assessment "go hand in hand" (p.51).

Airasian (1984) reviews literature that suggests teachers focus their classroom assessments in two areas: academic achievement and social behavior. The importance of these factors varies with grade level, with elementary teachers placing greater importance on social behavior. Airasian also found that teachers' informal "sizing up" assessments remain relatively stable throughout the year and influence student self-perceptions of ability.

Fleming and Chambers (1983), in a study that analyzed nearly 400 teacher-developed classroom tests, came to several conclusions:

- Short-answer questions are used most frequently.
- Essay questions are avoided, representing slightly more than 1% of test items.
- Matching items are used more than multiple choice or true false items.
- Most test questions, approximately 80%, sample knowledge of terms, facts, and rules and principles (94% for middle school teachers, 69% for high school teachers, and 69% of elementary school teachers).
- Few test items measure student ability to apply what they have learned.

Research by Carter (1984), in which the test development skills of high school teachers were studied, in support of what Fleming and Chambers found, reported that the teachers had

considerable difficulty recognizing or writing items that tapped "higher order" thinking skills, such as application. Stiggins and Conklin (1992), with a sample of thirty-six teachers, found that recall knowledge items were used approximately fifty percent of the time.

There is ample evidence to suggest that many teachers do not have sufficient knowledge and skill to develop, apply, and summarize classroom assessments. In a survey of 228 teachers from four grades (2, 5, 8, and 11), Stiggins and Conklin (1992) report that nearly three fourths of the teachers indicated some concern about their own tests. Examples of the kinds of concerns expressed included: "Are my tests effective? How can I make them better? Do they focus on students' real skills? Are they challenging enough? Do they aid in learning?" (p. 39). Concern was greatest for high school teachers. Only 15% of high school teachers indicated that they had no concerns about their assessments. Stiggins and Conklin also asked 24 teachers to keep a journal to reflect upon their assessment practices. The analysis focused on how teachers describe their assessments and what specific issues were raised related to their assessments. They found that teachers were most interested in assessing student mastery or achievement, and that performance assessment was used frequently. The nature of the assessments used in each class was coupled closely with the roles each teacher set for her students, teacher expectations, and the type of teacher-student interactions desired. The results of these investigations led to the development of classroom assessment profiles. The profile was tested with eight high school classrooms, resulting in the following key factors:

- Assessment purposes
- Assessment methods
- Criteria used in selecting assessment methods
- Quality of assessments
- Feedback to students
- Teacher as assessor (background, preparation)

- Teacher perception of the students
- The assessment-policy environment

These components can be used to characterize diverse assessment practices and environments.

Two recent studies document teacher beliefs and knowledge about classroom assessment. Frary, Cross, and Weber (1993) used a statewide random sample of 536 high school teachers of academic subjects to survey self-report practices and beliefs about classroom assessment. Frequency of use of various kinds of test questions revealed the following percentages:

Type of Question	Seldom or Never	Frequently or Always
Short answer	17%	56%
Essay	41%	38%
Multiple choice	21%	52%
True-false	47%	19%
Performance	30%	37%

These results suggest that teachers use a variety of assessment approaches. The teachers were asked to indicate degree of agreement to many statements concerning grading and assessment practices. Concerning assessment, it was noteworthy that 66% of the teachers agreed that essay tests provide a better assessment of student knowledge than do multiple choice tests; that 47% agreed that the nature of multiple choice items encourages superficial learning; and that better measurement occurs when teachers award partial credit rather than scoring simply right or wrong.

A second survey of teachers, taken in 1992, was structured to obtain teacher competency concerning assessment practices by asking teachers to indicate which of several possible answers to assessment questions was best (Plake and Impara, 1997). A national random sample of 555 elementary, middle, and high school teachers was used. Overall mean performance on the survey

was 66% correct. Teachers did better on items related to choosing and administering assessments and significantly worse on communicating results. According to the authors, the results "give empirical evidence of the anticipated woefully low levels of assessment competency for teachers" (p.67). The results also showed that teachers who had had a measurement course performed better than teachers who lacked this background.

In summary, the small amount of existing literature on classroom assessment practices indicates that teachers probably need further training to improve the quality of the assessments that are used. There continues to be reliance on selected-response tests, with conflicting evidence concerning the use of essays. Whatever the type of question, few are written to tap students' higher level thinking skills. Appropriately, teachers appear to use a variety of assessment methods. There is clearly a need for more research on classroom assessments. Classroom assessments consume significant amounts of time for both teachers and students, and have important consequences. Particularly absent in the literature are examination of relationships between classroom assessment practices and grading, how teachers use assessments to set standards, and how teachers make decisions about the assessments they use.

Teachers' grading practices have received far more attention in the literature than have assessment practices. This may be due to the salient and summative nature of grades to students and parents. Grades have important consequences and communicate student progress to parents.

A study by Stiggins, Frisbie, and Griswold (1989) set the stage for research on grading by providing an analysis of current grading practices as related to recommendations of measurement specialists and newly established Standards for Teacher Competence in Educational Assessment of Students (American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990). In this study the authors interviewed and/or observed 15

teachers on 19 recommendations from the measurement literature. They found that teachers use a wide variety of approaches to grading, and that they wanted their grades to both fairly reflect student effort and achievement, as well as to motivate students. Contrary to recommended practice, it was found that teachers value student motivation and effort, and set different levels of expectation based on student ability.

Brookhart (1994) conducted a comprehensive review of literature on teachers' grading practices. Her review identified 19 studies completed since 1984. Seven studies investigated secondary school grading, 11 studies both elementary and secondary, and one study elementary teachers. Three general methods of study were identified: surveys in which teachers responded to questions concerning components included in grading, grade distributions, and attitudes toward grading issues; surveys in which teachers were asked to respond to grading scenarios, asking what they would do in various circumstances; and qualitative methods, including interviews, observation, and document analysis. Despite methodological and grade level differences, the findings from these studies are remarkably similar. This suggests that conclusions warranted from the research are generalizable. Taken together, Brookhart comes to the following conclusions:

- Teachers inform students of the components used in grading.
- Teachers try hard to be fair in grading.
- Measures of achievement, especially tests, are major contributors to grades.
- Student effort and ability are used widely as components of grades.
- Elementary teachers rely on more informal evidence and observation, while secondary teachers use paper and pencil achievement tests and other written evidence as major contributors.
- Teachers' grading practices vary considerably from one teacher to another, especially in perceived meaning and purpose of grades, and how nonachievement factors will be considered.
- Teachers' grading practices are not consistent with recommendations of measurement specialists, especially confounding effort with achievement.

In one study, Brookhart (1993) investigated the meaning teachers give to grades and the extent to which value judgments are used in assigning grades. The results indicated that low ability students who tried hard would be given a passing grade even if the numerical grade were failure, while working below ability level did not affect the numerical grade. That is, an average or above average student would get the grade earned, whereas a below average student gets a break if there is sufficient effort to justify it. Teachers were divided about how to factor in missing work. About half indicated that a zero should be given, even if that meant a failure for the semester. The remaining teachers would lower the grade but not to a failure. The teachers' written comments showed that they strived to be "fair" to students. Teachers also seemed to indicate that a grade was a form of payment to students for work completed. More comments indicated that grades were something students earned as compared to grades indicating academic achievement, as compensation for work completed. This suggests that teachers, either formally or informally, include conceptions of student effort in assigning grades. Because teachers are concerned with student motivation, self-esteem, and the social consequences of giving grades, using student achievement as the sole criteria for determining grades is rare. This is consistent with earlier work by Brookhart (1991), in which she pointed out that grading often consists of a "hodgepodge" of attitude, effort, and achievement.

Cross and Frary (1996) report similar findings concerning the "hodgepodge" nature of grades. They surveyed 310 middle and high school teachers of academic subjects in a single system. A teacher survey was used to describe grading practices and opinions regarding assessment and grading. Consistent with Brookhart, it was reported that 72% of the teachers raised the grades of low ability students. One-fourth of the teachers indicated that they raise grades for high effort

"fairly often." Almost 40% of the teachers indicated that student conduct and attitude were taken into consideration when assigning grades. Interestingly, a very high percentage of teachers agreed that effort and conduct should be reported separately from achievement. Over half of the teachers reported that class participation was rated as having a moderate or strong influence on grades.

An earlier statewide study by Frary, Cross, and Weber (1993), using the same teacher survey that was used by Cross and Frary (1996), found similar results. Percentages of teachers agreeing or tending to agree to the following statements illustrate this conclusion:

Item	Percentage
• A student's ability should be taken into consideration in awarding final grades.	66
• An exceptionally low or high degree of student effort should be recognized by adjustment of the final grade.	66
• The amount of knowledge a student *gains* over the instructional period should be taken into consideration in awarding the final grade.	85
• Laudatory or disruptive classroom behavior should be considered in determining final grades.	31
• The minimum passing score on a test should be based at least in part on the scores earned by students of marginal ability who have be been putting forth satisfactory effort.	64

Another recent study by Truog and Friedman (1996), further confirms the notion of hodgepodge grading. In their study the written grading policies of 53 high school teachers were analyzed in relation to grading practices recommended by measurement specialists, and a focus group of eight teachers was conducted to probe reasoning used by the teachers. The study was based on an earlier investigation by Stiggins, Frisbie, and Griswold (1989). Friedman and Manley (1991) also found that teachers routinely use ability, attitude, effort, participation, and other factors in addition to achievement when determining grades. Truog and Frieman (1996)

found that written policies were consistent with earlier studies of teacher beliefs and practice. Nine percent of the teachers included ability as a factor in determining grades, 17% included attitude, 9% included effort, 43% included attendance, and 32% included student behavior.

Another survey of 143 elementary and secondary school teachers conducted by Cizek, Fitzgerald and Rachor (1995) collected data on teachers' assessment-related practices. Results indicated that assessment practices "were highly variable and unpredictable from characteristics such as practice setting, gender, years of experience, grade level or familiarity with assessment policies in their school district" (p. 159). Furthermore, teachers generally use a variety of objective and subjective factors to maximize the likelihood that students obtain good grades. Overall, the authors concluded that "many teachers seemed to have individual assessment policies that reflected their own individualistic values and beliefs about teaching" (p.160). The authors argue that grades should be used in more meaningful ways to communicate about student performance.

In summary, the literature on grading strongly supports the notion that teachers believe it is important to combine nonachievement factors, such as effort, ability, and conduct, with student achievement, to determine grades. While the studies are clear in this conclusion, less is known about how teachers decide to weigh these nonachievement factors in determining grades. Also, many of the surveys and other approaches in previous studies have asked teachers about their beliefs or projected behavior based on scenarios. It is possible that actual grading practice may be different. Despite increased focus on assessment and teacher competence with respect to measurement and grading, there appears to be a continuing discrepancy between recommended practice and teacher beliefs about grading. Furthermore, while descriptions of grading practices

are plentiful, there is little research on the relationship between grading practices and student motivation and achievement.

The literature reviewed on the nature and effect of assessment and grading practices on student achievement has demonstrated that there is little empirical evidence of the specific effects of using particular assessments and grading procedures. This is due in part to the complex nature of teaching, and how assessment and grading are only a part of instruction. Assessment and grading continue to be a private activity, with considerable variation among teachers. While "newer" forms of assessment, such as performance-based and portfolio, are based on recent research on cognitive learning, the suggestions are based on theory and not empirical evidence. There are several studies which show that teachers engage in assessment and grading practices that are not consistent with what would be recommended by measurement "experts." For example, combining nonachievement factors like effort, ability, and conduct with student achievement to determine grades, as well as "hodgepodge" grading. While descriptions of grading practices are plentiful, there is little research on the relationship between grading practices and student motivation and achievement. One theoretical model postulated by Brookhart (1997) represents an initial perspective about how assessment and grading practices affect self-efficacy, effort, and achievement. There is a strong research base with respect to the two major contributors to motivation (self-efficacy and importance, utility, and value), but not much about how specific assessment and grading practices effect these two components.

Research Questions

The purpose of the proposed research (both phases) is to gather information from teachers regarding their assessment and grading practices to answer the following questions:

- What is the current state of assessment practice and grading by teachers?
- What assessment and grading topics do teachers identify as needs to be addressed in in-service?
- What is the relationship between assessment and grading practices and grades given to students?
- What are the relationships between grade level, and subject taught and assessment and grading practices?
- What are the reasons teachers give for their assessment and grading decision-making?
- What is the impact of the SOL tests on the extent to which different assessment techniques are used in the classroom?
- What classroom assessment and in-service needs do teachers have?

Forshawdowed research questions guiding Phase 2 include:

1. What is the nature of teacher decision making concerning classroom assessment and grading practices?
2. What influences teacher decision making concerning classroom assessment and grading practices?
3. What classroom assessment and grading practices are identified?
4. What justification do teachers give for their classroom assessment and grading practices?

Methodology

Research Design

The research consisted of two phases, one involving a written survey of a large number of teachers and one using face to face interviews. Phase 1 included development and administration of a teacher questionnaire to survey teachers' assessment and grading practices and in-service needs. Quantitative analysis of the data included data reduction, descriptive statistical results, and the investigation of relationships with analysis of variance and correlational procedures. Phase 2 used interviews with selected teachers to investigate decision-making and justification for specific assessment and grading practices. This report is concerned with Phase 2 of the research. A qualitative research design was utilized.

Participants

On the written survey from Phase 1 teachers were given the option to participate in an in-depth interview regarding their responses to the survey. Volunteers' surveys were pulled and reviewed for maximum variation in item response. Surveys in which maximum variation responses were consistently observed were selected for further interviews. Maximum variation responses were identified via the survey Likert scale in which a response of 1 indicated "not at all" and a response of 6 indicated "extensively." Sixty (60) surveys were originally identified as meeting the maximum variation criteria for selection. The 60 teachers were then contacted by telephone and a letter was faxed to their school requesting an interview date and time. Of the 60 teachers originally selected for an interview, 28 ultimately participated in the interview process.

Of the teachers interviewed for this study, English/language arts classes were represented by teachers from more than a dozen different schools in 7 different school districts. Grades represented by the teachers were 5, 6, 7, 8 and 12, with students of varying academic abilities (low, moderate to advanced placement).

Math classes were also represented by teachers from more than a dozen different schools in 7 different school districts. Grades represented by the teachers were 7, 8, 9, 10, 11 and 12 with students of varying academic abilities (low, moderate, to advanced placement as well.

Table 1 shows the breakout of teachers according to subject matter, grade level and ability level.

Table 1
Summary of Characteristics of Teachers Interviewed

	Grade Level							
	5	6	7	8	9	10	11	12
Eng/LA Ability								
Low		1						
Mod	1	1	1					
High		1		2				4
Math Ability								
Low			2	2	1			
Mod				2	1	1		
High			1		1		2	2

Interview Protocol and Process

Four members of the research team participated in the interviewing process. All but four of the interviews were tape-recorded. Interviewer notes were taken on the four unrecorded interviews, as well as for some of the tape-recorded interviews. Interviews lasted 45 to 60 minutes and took place in the teachers' schools. An interview guide was developed by the research team prior to the interviewing process. It was used in four pilot interviews and revised by the team prior to completing the sample of 28 interviews. A copy of the interview guide is attached in Appendix A.

Data Analysis

Each tape recorded interview was transcribed into approximately 8 to 14 pages of single spaced typed text. The text was then loaded into HyperRESEARCH qualitative software. Data were reviewed by team members and then coded by the university team members according to the emerging topics of the interviewees, as well as the pre-established topics identified in the interview guide. Forty nine (49) codes were initially established, although 26 became the major categories used throughout the data analysis. Following coding, members reconvened to review the coding and categorizing. The categories were then further collapsed into four themes that explain the data. Table 2 shows the initial codes identified during the data analysis, as well as their level of frequency throughout the data (codes with an asterisk [*] beside them indicate major categories).

Table 2

Summary of Codes and Frequencies of Responses

Code	Frequency
Administrative	1
Advice to other teachers	2
***Assessment rationale**	53
Assessments drive lesson plans	9
***Borderline grades**	18
***Choice of assessments**	107
Class type	34
***District grading policy**	19
***Effort versus ability**	24
***Extra credit**	21
***Feedback**	46
***Formal versus informal assessments**	11
***Grade challenged by parent**	2
***Grade distributions**	36
***Grades**	29
***Grading policy**	50
***Grading policy rationale**	25
Group work	8
***Homework**	4
***Influences on mode of teaching**	17
***Informativeness of assessments**	30
Lesson plans	2
***Lesson plans drive assessment**	27
***Modification of assessments**	13
***Modify assessments for spec. students**	18
Objective assessments	21
***Ongoing assessments**	9
***Other teacher grading policy**	12
Performance assessments	1
Pre-assessments	40
Publisher made assessments	8
***Pulling for students**	39
Quiz grades	6
Revision of assessments	23
Socio-economic status	1
***SOLs**	52
Standardized tests	2
Student effort performance	22
***Student motivation**	57
Student performance on quizzes	19
Student recall knowledge	21
Summative evaluation	1
Teacher made tests	7
Timing of assessments	7
Types of test items	2
When assessments used	2
***Worth and value of assessments**	6
***Zeros**	22

The selection of codes into categories was not based on frequency counts alone, but also depended upon the similarity a code had to other codes which were becoming major categories. For example, the code "grade challenged by parent" has a frequency count of only 2, yet it was still retained as a category because of its similarity to other "grade" related codes and categories.

Throughout the process of coding and categorizing, an on-going inductive analysis of the data was also occurring. The researches were constantly interpreting the data as it unfolded in an effort to understand themes as they emerged. During this process of inductive analysis, coding and categorizing decisions were also affected such that codes and categories that initially promised explanatory value for the data were eventually discarded as new, more meaningful codes and categories developed over time. For example, the code "student recall knowledge" has a frequency count of 21, yet it did not become a major category because the on-going inductive analysis revealed, over time, that the code did not hold much explanatory value for the data.

Following the reduction of 49 codes down to 26 categories, the categories were further reduced to four (4) themes: (1.) Teacher beliefs and values, (2.) External Factors, (3.) Teacher decision-making rationale, (4.) Assessment and grading practices. These four (4) themes have powerful explanatory value for the data and offer a tentative theory of teachers' assessment and grading practices.

As a validity check, 50% of the coded transcripts were peer reviewed by a MERC associate to determine agreement on the selection of codes assigned to chunks of data. Of 520 codes assigned to the data by the researcher, the peer reviewer agreed with 450 of them. This resulted in an 87% rate of agreement between the researcher and the peer reviewer.

Findings

The results will be presented by first explaining an overall model of teacher decision making that represents a synthesis of external factors that influence teachers and teachers' beliefs and values. Following the model, quotes from teachers that represent their beliefs and values, external factors and their decision making rationales will be presented.

A Model of Factors Influencing Teacher Assessment and Grading Practices Decision-Making

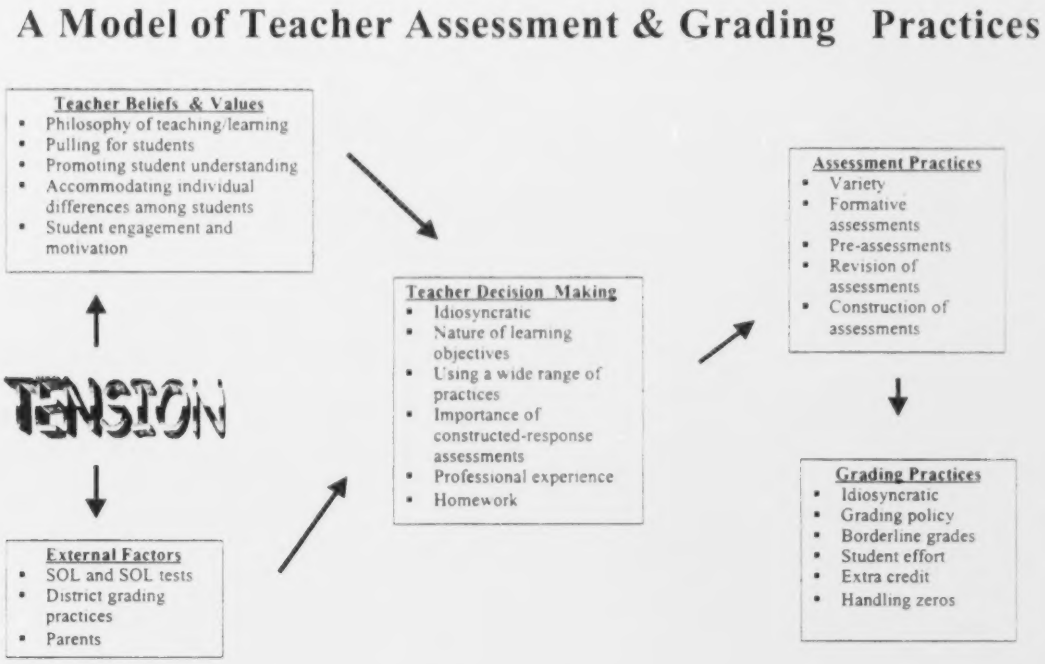
Inductive data analysis resulted in a tentative model that explains how and why teachers decide to use specific assessment and grading practices. The main tenet of the model holds that there is tension between internal, beliefs and values of teachers and external factors that are imposed on them. This tension between these two types of influences is apparent in the explanations teachers give for their assessment and grading practices. Such practices are influenced most heavily by internal beliefs and values. External pressures, especially recent SOL testing, forces teachers to use assessment and grading practices that probably would otherwise have little impact. Greater tension arises when external pressures increase, and lessens as teachers gain experience.

Within this model external factors are considered to be accountable (to systems, parents and students), with their end result revealing an "objective" way of documenting student performance. Teachers' beliefs and values, on the other hand, are internalized factors that are frequently idiosyncratic beliefs that comprise teachers' philosophy of education. They usually consist of beliefs that students do not fit into highly organized, objective, categories, but rather are individuals in need of flexible assessment and grading schemes. As a result, teacher

decision-making rationales are varied, individualized justifications for the types of assessments used and the grade assigned to a student.

Figure 1 illustrates the model of teacher decision making. It shows how the decision making process is influenced by teacher beliefs and external factors, leading to rationales for specific assessment practices, followed by grading practices. The model focuses on the decision making process, but both teacher beliefs and values, and external pressures, will influence assessment and grading practices in a fairly direct way. In the end, to better understand teacher decision making, it is necessary to explore factors in each of the two major categories of influences.

Figure 1. A Model of Teacher Assessment and Grading Practices Decision-Making.



Teacher Beliefs and Values

The most salient internal factor that appears to influence teacher decision-making concerning classroom assessment and grading practices is the teacher's philosophy of teaching and learning that provides justification for the practices. This internal set of values is important because it provides a rationale for using assessment and grading practices that are most consistent with what the teacher believes is most important in the teaching/learning process. For some of the teachers interviewed the philosophy of learning was focused on doing whatever was needed to help students succeed, to "pull for students." In extreme cases, this meant significant modifications of assessments, such as writing multiple forms of tests to accommodate different students' needs and abilities, allowing creative expressions such as artwork to substitute for

BEST COPY AVAILABLE

regular paper and pencil tests, and allowing students to veto certain types of tests questions if they feel incapable of responding to them. Other teachers were less accommodating but still indicated a philosophy based on student success. These teachers, for example, would accept late work or revisions of work.

Essentially, it's as if assessment and grading practices are whatever will best serve the purposes that are linked to a larger, more encompassing philosophy of education. For example, teachers believe that students need to be meaningfully engaged in learning, and would use assessments and grading factors that would enhance this engagement. Five categories of teacher beliefs and values were identified: Philosophy of teaching/learning, pulling for students, promoting student understanding, accommodating individual differences, and student engagement and motivation.

Philosophy of Teaching/Learning

Note how the following excerpts from the teachers frame their assessment and grading practices in a larger philosophy of teaching/learning:

- I weigh more on homework …there are worksheets that I use. And to me, my philosophy of education is run by Dewey. The more you practice something the better, the more proficient you become in that skill.
- The daily grades are my way of making sure that they continue this process like it's supposed to be. If they haven't done the daily things that I've asked them to do then they are not going to be able to do that end result.
- I always assess early on to see what people know so that I could split groups and remediate or accelerate as needed.
- To me grades are extremely secondary to the whole process of what we do. I have goals to what I want to teach and I use assessment so that I know what I need to work on. What people have mastered and what they haven't.

Pulling For Students

For most of the teachers it was evident that they wanted very much for students to succeed and to obtain good grades. We have labeled this value "pulling for students." It is manifest in assessment and grading practices that are designed to give students the best opportunity to be successful. In some cases it almost appeared as if teachers were using specific practices so students could pull up low test scores. Working within the constraints of the grading system, teachers wanted very much for students to do well. Here are some illustrations of this tendency:

- Maybe this is the dedicated teacher's syndrome or whatever. I'll chase the kid around for a long time so I can get a few points.
- I'm always trying to find some ways so that all the children can find success, not just Johnny and Suzy getting the A but also Sally and Jim can get an A also.
- Everybody takes the quiz but the way I record the grade is only the good grades. If you don't get a B or better then I don't record it.
- I also try to give opportunities and assess them in different types of ways and give them the opportunity to, if they blew something, … I give them the opportunity to make up points or get bonuses.
- When we do tests and quizzes I'll divide them into sections so each of my learners will have at least one area on the quiz or test where they'll be able to shine.
- You want to have a variety of activities, because they may shine in one and not in the other.
- I always found that the more ways you can assess children and the more grades you have for them, if they have failed a couple of tests or couple of projects or missed a couple of homeworks, you can take that out and still keep their grade up because you have a large quantity.
- I never tell them this but when I do that average I'll add and extra 5 points to everybody, and as far a borderline is concerned, most of the time I round up.
- When I have kid that is between a D and an F, I go back and really scrutinize [so see] if there is something maybe I did or left out or didn't … I'll go back and re-grade something.
- I always tell them that they are here to succeed, not to fail, and my classroom's designed for them to succeed.

Promoting Student Understanding

An important component of the philosophy of teaching/learning is to gauge student progress by using assessment to check for student understanding. This came up many times in the

responses of the teachers. They are very concerned about getting students to the point where they truly comprehend and understand, not merely memorize.

- It's not only me lecturing and then they soak it in and then regurgitate back to me. I always tell them that I could train a monkey to do that if I give them enough bananas. That's not what education is, that's not my goal here. My goal is that you can understand it, so you need to participate.
- And to me that's how you approach math. It's not memorizing because first of all you're not going to retain it and that's not going to help you.
- You want to know, what have they really learned or can they apply it … to get a more realistic grade of what the student really does know about the material.
- So we have to read it in class [Macbeth] … then I say, 'alright, this is what I'm going to test you on … I give them samples … then I decide who gets what kind of quiz. It's a lot of work, but the thing is it makes them more successful.
- The assessments where students actually have to show me some work or write about are most valuable for informing me about how much students know. Because it's then that you know that they understood every process that tells you more about a student than just grading a sheet of answers.
- Big tests and essays best because this is where higher levels of thinking come in.
- I go back to the ultimate, I don't care how I get them there, I want them to learn it. And if it means I will give you 2 more points for this if you go back and fix it and get it right … if I have to dangle that carrot to get them three … I'll get them there.

Accommodating Individual Differences Among Students

Another aspect of assessment that appeared linked to philosophy was varying assessments to accommodate individual styles. Most of the teachers made efforts to provide varied assessment that meet a variety of learning styles. This is, again, part of a larger philosophy of education – that individual differences among students are important and need to be considered – in all aspects of teaching, including assessment and grading. What teachers do is essentially modify assessments on the basis of student characteristics. Following are some illustrative quotes:

- I think it'll go back to the goal I have: try to meet the needs, interests, capabilities of the children. If you don't have a variety of things, you're not really focusing on what that child's ability is.

- I will have at least two versions of any given test or assessment … there are always three versions and most often four depending on who I have in my class … and I think it is very important to do it … I always have several ways [of assessing].
- Some students do better on paper than they do orally, some students do terrible on paper but you know that they know it so you have to come up with a way to say show me what you know. My philosophy is I'm trying to get them to show me what they know, not trick them into showing me what they don't know.
- I feel that kids learn in all different ways, and they have different ways of showing it.
- They [assessments] change based on the needs or capabilities of the students. I'll make some tests easier and some harder, depending on the ability level of the student.
- We tend to always have a fairly big disparity in student ability … there's no use in teaching things that people already know, and there's no use to teaching way over people's heads.
- The types of quizzes do vary .. I try to accommodate. I wouldn't put as many formal proofs on their quizzes or tests as I would with an honors class.
- It's really, really important that you know the kids individually as people and you have to know their stories.

Student Engagement and Motivation

Teachers clearly indicated that it is imperative for students to be actively engaged in learning, and, hopefully, motivated to do their best work. This engagement and motivation is seen as critical to the learning process. Consequently, teachers base their assessment and grading practices decisions on what will result in the greatest amount of student engagement and motivation. As will be pointed out later, this results in using many constructed response assessment items and a fairly heavy emphasis on homework. Here are some quotes from teachers that illustrate the importance of engagement and motivation:

- If you really want a student to learn, the student has to be actively engaged … and doing group work … I find that works best.
- Everybody has to be involved in this, not just those who look like they are falling asleep, but everybody … we'll continue until everybody has their chance.
- The reason I do it [use a mix of assessments] is because I want my children to be task oriented, I want them to be responsible for every assignment a teacher gives. I don't want them to think they can skip an assignment.
- Students learn more when students are actively engaged. Daily grades are based on participation in groups.
- Students need to be task oriented and organized.

- It's worth a few extra points in their grade because it means that everybody in the whole school triangle [student, parents, teacher] is involved in their education.
- Use essays to get them engaged, to motivate them.
- I use a goody jar … it's not really assessment … but really helps me in my assessment, especially with kids with low motivation.
- I make them do something. Make them learn and see that they have to put forth some effort. They need to know that they have earned what they've got.

External Factors

External factors are influences that originate outside of the classroom. They are not under the control of the teacher, but still impact the nature of classroom assessments and grading practices. Three major external factors were identified: Standards of Learning (SOL) and SOL tests, district grading policies, and parents.

SOL and SOL Tests

This category was included on the interview guide rather than inductively derived from the data. Since it is clear that the SOL and SOL tests have had a great influence on teachers in general, the intent here was to focus the teachers' thinking on how the SOL tests affected their classroom assessment and grading practices. The teacher comments indicated that the SOL were, in fact, impacting their classroom assessments and grading. This was typically not a radical or far-reaching influence. Rather, the SOL tests provided an "external" reason to modify their classroom assessments so that they covered more of the SOL, and, to a lesser extent, so that they used the multiple choice question format to a greater extent. Using more multiple choice questions would better prepare students for the SOL tests. Teachers seemed to feel resigned to making these changes and seemed to suggest that without the SOL tests they probably would not have made the changes. The comments below capture these perceptions of the teachers. In the

comments it is evident that there is tension between what the SOL and SOL tests suggest should be assessed, and how, and the assessment approaches of teachers based on their own beliefs.

- I use teacher made tests because I feel like I'm the one who has taught it and I know what I'm looking for. But as the same time I'm going to make it up according to the SOL that we have to follow.
- What they did was they [SOL tests] defined it [classroom assessments] ... in math most of them are not multiple choice tests, but I give them more multiple choice so they can get used to it ... with a multiple choice test I don't think you get an accurate evaluation of the students' knowledge.
- I think the teacher has to teach both for the SOL test, which is a necessity today, but you can't forget a lot of other things ... you have to have a balance of both.
- I do it sporadically so it's familiar to them, but it's not my general way of doing things.
- As far as changing my grading practices, probably it will change my assessments somewhat. I've got to make myself do more multiple choice questions.
- We're testing more often, we're giving samples of standardized tests, multiple choice tests.
- On my wall over there is SOLs and I know exactly what's got to be on that test because we have practice SOL tests that we've been given to preassess the students. So assessments are beginning to drive my lesson plans.
- Assessments are now not just to know (what students know), but to prepare them to have test taking skills.
- It is a good thing in terms of having them ready for standardized testing.

The following comments show how some teachers have changed their classroom assessments to conform to the SOL, but do not believe this is in the best interest of the students. The impact is undesirable because it means content not on the SOL tests is much less likely to be assessed or emphasized. Here assessments are driving instruction, and to the extent that the classroom assessments are influenced by the SOL and SOL tests, the greater this external influence is on teacher practices. Conflict also arises between assessments that teachers believe give them greater understanding of student knowledge, typically constructed response items, and multiple choice tests, which are viewed as limited in what they tell the teacher.

- I think you're doing the children such a horrible disservice when you teach the

SOL tests because you leave out so much wonderful stuff that some of these children will never get anywhere else.

- I am opposed to the SOL testing, it just doesn't leave any room for individualization on the part of the teacher.
- They [multiple choice tests] don't always tell me what I want to know.
- I like constructed answers because you can go back and show the child where they messed up, I guess multiple choice measures it the child knows or that the child is a good test taker.
- You know frankly, with few exceptions, if we didn't have to get ready for standardized tests I'd probably seldom use multiple choice as a format.
- I'll tell you that this year has been very different from the years past because I always gave tests and quizzes that required show me the work. I didn't give multiple choice tests. Because the students now are required to take these SOL tests, now I'm mixing … some multiple choice and some show me the work … yea, outside forces control me.
- This year with the SOL coming in a lot of revision is needed.

District Grading Policies

Each teacher was asked about the effect of division grading policies on their grading practices. While each division has such a policy, it was evident that teachers use such policies only in a very general way, and that their own approaches and preferences are much more important. This contributes to greater diversity of grading practices among teachers. In some cases teachers completely ignored division policies.

- Fifty percent or less [is driven by the school or division policy].
- It's my decision as to how I interpret school policy.
- I am somewhat compelled to go with our numerical system with the county.
- I go with my own judgment also; a little bit of both to be totally honest.
- We got one [division policy] this year. I was furious about it. I'm finally getting some things right after 30 years, and they told me I couldn't do things ... it's not the grading, it's the process of learning.
- I think they [division] are more concerned with you having enough grades every 9 weeks.

Parents

This category is one that was derived inductively from the data. There was little parental influence on the nature of assessments, but clearly teachers were influenced by parents in the grading system they use. Teachers want to be able to meet with parents and provide reasonable

explanations for the grades they have given. The most important factor is having sufficient justification for grades to avoid parental conflicts. For some teachers this meant using very specific, "objective" scores and averaging, while for most teachers there needed to be a sufficient number of grades to show clear patterns.

- To me the calculator is the deciding factor ... I can sit down with any parent or any kid ... it makes my life easy, because I punch the buttons, hit the equal sign, and there it is.
- When you're going to show Johnny's mom whey he got a B, you won't have a lot of reasons to show his mom if you only have one test out a nine weeks.
- Parents were a bit more prone to suing teachers ... as a result we had to develop our own objective ways of assessing students' academic performance.
- I don't give A students a C without their parents knowing ... challenges come from a lack of communication between the parents and teachers ... I bring parents into the process on the first day ... talk to my parents all the time.
- I'm very up front with them on how I do report card grades ... I tell them that when they get report cards they will be based on objectives and how well students met them.
- Pretty iron clad, does not leave any room for any subjective evaluation at all. It saves a whole lot of arguments with parents.

Teacher Decision-Making

Teachers were asked repeatedly to provide a rationale or justification for *why* they made assessment and grading practices decisions (e.g., why the types of assessments used, why specific factors used in grading). Overall there was great difficulty and some uneasiness in the responses of the teachers. It was something of a struggle for them to explain, particularly if they had been teaching for many years. One general assumption seemed to provide a foundation or basis for their rationales. It was apparent from the interviews that teacher decision-making is a highly individualized, idiosyncratic process. Thus, no two teachers were alike, and the comments suggested that they believed they should not use the same assessment and grading practices as other teachers. Furthermore, with some probing, five additional factors emerged as significant in identifying the reasons teachers give for their assessment and grading decisions:

the nature of learning objectives, using a wide range of practices, the importance of constructed response assessments, professional experience, and homework.

Nature of Learning Objectives

Many of the teachers indicated that the nature of the learning objective would determine the choice of assessment method. Simple recall knowledge, emphasized through drill and memorization, would be assessed used selected response items, such as multiple choice or true/false, while objectives emphasizing thinking skills, such as application and reasoning, would be assessed with constructed response assessments, such as essays and performance assessments. The following excerpts show the influence of objectives and topics.

- Well it depends on the topic sometimes, for example, I just finished the objective on surface area and volume and in that case I do a lot of handouts and worksheets. With learning definitions I'll do matching, multiple choice type items. So a lot of times the topic will determine the assessment that goes with it.
- The manner that I assess tends to be more related to the subject matter that I am teaching … with teaching grammar, just multiple choice items type things for that because there is some memorization involved.
- I do some grammar quizzes, we are going to have a grammar test, they're really sporadic because it really has to be revealed in their writing.
- It depended on what were covering as to how we needed to assess them.
- Pop quizzes that they don't know about that I usually try to give them to get an idea of whether or not they understand the material.
- I try to assess at the point that I feel pretty confident that the children understand the material. That's the point at which I assess.

Using A Wide Range of Practices

A second finding was that all the teachers believed that they should use a variety of assessment methods and should use multiple criteria in grading. This may reflect the conflicting influences of internal and external factors, but probably is based mostly on the belief that since no two students learn the same, multiple assessment methods are needed to fairly assess them so

that all students are able to demonstrate what they have learned. It's also consistent with the "pulling for students" belief, use whatever assessment best matches with student styles and strengths to give the best performance. Notice in the following comments how assessment practices are both varied and influenced by the nature of the students.

- It's what I feel like the kids in that particular group, how I'm going to find out what they know in the best way … some students do better on paper than they do orally, some students do terribly on paper but you know that they know it so you have to come up with a way to say show me what you know. My philosophy is I'm trying to get them to show me what they know.
- I use a little of everything.
- Day to day, observation, the almighty observation. You're listening, oral presentations, looking at a project, at a test score, it's everything that a child can give you.
- They [assessment practices] change based on the needs or capabilities of the students. I'll make some tests easier and some harder depending on the ability level of the students.
- You have to adjust [assessment practices] to where you are.
- I tend to rotate they types of assessments so that they have a lot of different types.
- I try to give opportunities and assess them in different ways and give them the opportunity to … if they blew something, in a projects format it seems like its easier to assess … I give them the opportunity to make up points.

Importance of Constructed Response Assessments

It was clear that the vast majority of teachers preferred constructed response assessments, where students "show their work" (e.g., short answer, essay, performance assessments, demonstrations, exhibitions, portfolios). The teachers indicated that these kinds of assessments give them the best indication of whether students truly understand and can apply what they have learned. This is consistent with the internal belief that assessments should serve instruction by showing what students understand. Caution was indicated in the extent to which objective assessments can provide sufficient evidence that students actually understand as compared to memorization. Here are some illustrative comments:

- Whereby I use rubrics to score a lot of their projects, I also try to do as much hands on as possible.

- Observations, rubrics, anything that will show you a measurement of a child's performance level.
- I like the rubrics. Just because it allows for more creativity on the child's part. It seems like they are giving me more information.
- One project I like is called shop till you drop. They're required to go out and comparison shop, try to find the same items on sale at three different stores … they have to do different things to get it … so I use that as a test grade.
- I use written open format evaluation. I occasionally use matching and multiple choice. I do a lot of personal anecdotal evaluations. In other words, live performance tests.
- I might start with a quiz and then if it was still unclear then I would go to a personal one-on-one oral assessment or a task assessment … a project type of assessment.
- When it comes to science it would tend to be more free form drawing open-ended questions some task oriented things where they have to do something … the same goes with social studies … more open-ended assessments too with some hard knowledge types of stuff.
- I have some multiple choice, personally I do not like those, I would rather have free response because then they have to put down exactly what they know.
- I always use rubrics … and when we do book reports, or any product based outcome, when we're building or making something … we use the rubric.
- My tests are for the most part essay … generally really thorough and quite long … all the kids have to write three essays.
- If I teach geography, the way I would assess that is to give them a blank map.

Professional Experience

Teachers' experiences have evidently had much to do with determining their assessment and grading practices. Whether by trial and error, or by talking with others, it seems that the teachers learned through their own experience which assessments and grading approaches would work best for them and their students. It is as if the practices simply evolved over time. One thing that was absent in their comments was any indication of influence from either initial teacher training or subsequent professional development opportunities. The following comments illustrate the importance of experience:

- I've taught for twenty some years and I guess some of this just evolves over the years.
- I had to figure out what to do. Sometimes you talk with other teachers and find that they are doing different things, but I don't know that I have talked to any other teachers who are doing what I do. It sort of came upon me … trial and error would be the best answer,

which would put it all in a nutshell … like a lot of things, once you do it for a long time you sorta get a feel for it.
- Test experience, what was done with me in high school.

Homework

Finally, one common thread among most teachers was the importance of homework. It was indicated that homework, much like quizzes, provide the teacher with an immediate indication of student understanding. Homework also is important to student learning. Most teachers believed that homework was essential practice in the skills. In this sense, then, homework provides an added learning activity and an indication of understanding. Here instruction and assessment are integrated. Here are a few comments indicating the importance of homework:

- I weight [grades] more on homework, say around 40%.
- I found out that if a student starts homework in class … [put] homework assignments on the board and go over them … you ought to be able to help them.
- I give what I call a mini quiz every day. That's o the last night's homework.
- I'll teach a lesson, they'll have a homework assignment, they bring it in the next day and we'll check over it and I'll check and see who has it because use that way then you get something wrong on the homework, I expect them to ask questions.

Assessment Practices

Based on beliefs and values, and on external influences, teachers select and implement specific assessment practices. The variety of different types of assessments used reflects teacher beliefs that informal, observational assessments and constructed response assessment are best for gauging student understand on the one hand, while external pressures tend to result in more objective items. As a result, most teachers use the same variety of assessments, individualized to their students and based on their own experience and the nature of the learning objectives. This would include homework, quizzes, tests, performance assessments, and participation. Using different kinds of assessments also allows more students to show their best work. Several

themes concerning assessment practices emerged from the data, including formative assessments, pre-assessments, revisions of assessments, and construction of assessments.

Formative Assessments

From the responses of the teachers it became evident that some assessments are more informative than others. The daily checks and observations, what might be called formative or informal assessment, were clearly most informative for the teachers. This kind of assessment is ongoing and continually informs instructional decision-making. Here are some examples of what they said about the informativeness of assessments:

- My informal assessment is [most informative].
- Daily quizzes. Yes, especially with daily quizzes as a check of previous day and use quiz to go over concepts as needed.
- Ones where students apply what they've learned [are most informative].
- Daily quizzes [are most informative] … gives daily pulse of learning for the teacher.
- Quiz, graded by both teacher and student [is most informative]. Gives a quick overview of class progress.
- Well, definitely the free response over the multiple choice … sometimes the group assignments I give … also listening to them talking in groups.
- Using a rubric with very specific guidelines [is most informative] for me and them [students].
- Probably I can find out a whole lot more in oral. Asking them to explain something to me … and just watching them on a day-to-day basis.
- Probably the tests. Because like I said, you can pretty much realize that they have been able to master larger chunks of information that's opposed to isolated things.
- Almost always for my purpose it's the writing assignments that I spend in the rubric [that is most informative].
- Probably quizzes on 2 or 3 sections [are most informative].
- Class work assignments [are most informative] because they are one in class under supervision.
- Projects most important, also papers.
- Oral questions used extensively and daily assessments.
- Homework and class participation.
- I develop a rubric for every writing assignment and I show the children that rubric in advance as well. … it lists the skills I'm looking for … it's a diagnostic kind of thing … it gives me the ability to cover things they've learned in the past and weak spots.

Pre-Assessments

Another area that was brought up in the interviews was the nature and use of pre-assessments. These are assessments done prior to instruction or beginning a unit. It was clear that most teachers use some kind of pre-assessment. This was usually in the form of an informal review of current student knowledge, understanding, and skill, done through classroom observation, short quizzes, and through question/answer sessions. However, some teachers actually gave formal pretests. Also, some teachers interpreted "pre-assessment" to mean "expectations," which they tried not to make. Finally, pre-assessments are affected by subject matter and experience. In highly sequenced subjects, such as secondary mathematics, there is less pre-assessment. The more experience a teacher has, the less likely he or she will use pre-assessments.

Here is a representative sample of comments of teachers regarding pre-assessments:

- I'd rather judge on what they are doing, not on a standardized test.
- I do pre and post-tests three or four times a year … if you don't do your pretest in the first or second week of school then you can just hang it up.
- If you preassess then you're going to be able to plan better and you find out student needs.
- I use writing as pretty much a gauge to find out what the students are lacking in and what they need to know.
- I also give them a pre-assessment.
- I always assess early on to see what people know so that I could split groups basically and remediate or accelerate as needed. I am a real stickler to assessing only to find out what people know.
- What I will do is I will kind of explore their knowledge so to speak. I might start with a quiz and then if it was still unclear then I would do a personal one-on-one oral assessment.
- At the beginning of the year I'll give them a math test o the SOLs and let them see what they couldn't do.
- Sometimes I will use a diagnostic test … some of the kids you think are very articulate but that doesn't mean they know some of it.
- Yes, with daily quizzes, especially with advanced students who need to move ahead of the rest of the class.

- Informally, usually in beginning a unit.
- Not so much in algebra … mostly new material … find pre-assessing discouraging because so much is new.
- Pretest to find location that class is in to start the year.
- Not as much as I probably should … I used to give a pre-assessment of grammar just to see where they were … they were all over the place.
- Sometimes we do a little bit of that [pre-assessment]. I usually have a pretest, maybe the first day to kind of see where they stand. I used to do it more than I do now.
- We have a little survey we do.
- Sometimes. There are only four math teachers here and I know the 6th grade teacher and what she covers … I only teach those things they don't know.
- Yea, because at the beginning of the year for the math class there is a lot of review.
- No, because they have been pre-assessed because of the placement into the pre-algebra group.
- Yes, informal, with short quizzes or examples.
- Yes, informally with questions and answers at the board.

Revisions of Assessments

Teachers were asked to comment on the extent to which they revise assessments, when the revisions would be done, and the nature of the revisions. Almost all of the teachers indicated that tests and other assessments are revised both from year to year and from one week to the next as the testing date approaches. This constant revision process is done because students change from one year to the next, because the content of what is covered changes somewhat, and because in each class there are special circumstances that affect what should be tested. This further supports the teachers' need to adjust assessments to individual differences of students and to the objectives being covered, all in the goal of pursuing increased student learning. It also points to a significant time commitment for teachers.

- Every year it's different … I don't think I've ever used the same test two years in a row on anything.
- The only thing that has stayed the same is the spelling quizzes … I change everything else … I keep a copy of it so I can see the change.
- You modify everything. If everybody fails the test then I modify it because I've done something wrong … I try to write and revise tests students take within the next two days.

- Why are they revised? Because the results that were found on previous tests were not satisfactory, did not show student performance.
- I usually change them pretty much each year.
- I rewrite them every year, maybe not entirely, I'll use some parts.
- When I grade a set of papers and there is something there that the children are not understanding, I go back and revise the assessment … so there's a constant revision process going on.
- It [revision] gets done on a regular basis … this year I've modified almost all of my tests.
- If I find an old quiz, it just doesn't work out for them [students] … it's the wrong thing. They're not there yet or they're way behind it or what have you … it [using old tests again] just doesn't work.
- The weekly tests, you're constantly changing based on the needs of the class.
- Yes, I do go back and modify them [tests], I don't just pull it out of the cabinet and give it to them.
- I don't even look from one year to the next to see, I always rewrite them. I know from last year to this year they are a whole lot different. I don't just recy7cle.

Construction of Assessments

Teachers may use assessments they themselves construct, they may use tests provided by publishers or school divisions, or they may use some combination of these, each influencing the other. Overall, these teachers clearly rely most on teacher-made tests, ones they construct. Publisher's tests are not widely used because they do not address local contextual factors such as what was covered and the characteristics of the students. Note how the following quotes emphasize the importance of teacher-made assessments:

- Over the whole year generally, I do teacher-made tests.
- I use teacher–made tests because I feel like I'm the one who has taught it and I know what I'm looking for.
- Sometimes I will pull questions from a pre-made test but I don't generally like to give an entirely text book made test … I don't tend to teach things like they are presented in the book, so I make them [tests] up.
- Some of them I create myself, some of them are from the text. When I take them from the text, I very rarely give the whole thing, I usually do bits and pieces and kind of paste and put together.

Grading Practices

Grading practices represent an interesting mix of results from assessments and deciding how to weight different factors different amounts. In addition, there clearly are external factors, such as division grading policies and parents, as well as teacher beliefs about motivation and engagement that influence the practices. What results is, like the nature of assessments, individualized approaches that take these considerations into account with the types of students in the class. In discussing grades several factors emerged as significant, including grading policy, borderline grades, how effort is handled, how extra credit is handled, grade distributions, and how zeros are handled.

Grading Policy

Regardless of division or school policy, teachers have their own grading policy. And it seems that most teachers have unique or idiosyncratic procedures. However there are some common elements. For one, all teachers obtain many grades from primarily four sources, homework, quizzes, tests, and projects or papers. Some teachers also utilize participation, class work, or some other indicator of effort. Interestingly, tests typically do not account for more than 30% of the final grade. Teachers indicated that they use a criterion-referenced approach to grading rather than a norm-referenced approach, and typically would use a total point system that provides percentages consistent with division guidelines. An interesting issue is whether the teacher uses students in the class or grade level objectives as a frame of reference for giving grades. That is, it is possible for students to receive As if they learn a lot, or receive Cs for the same level of performance if what they have achieved is below grade level.

- I rely on tests only 30%; class work 65%.

- I'm not a believer in having a bell shaped curve for grades … if in a class nobody's trying and I only have one or two Bs and the rest are Fs, that's exactly what the assessments are going to be.
- If I have students who are working on the first grade level they necessarily get Ds and Fs on the report card where I'm grading 3rd grade objectives.
- They give me objectives … I always to way , way beyond those.
- They have 13 grades, I drop the 3 lowest … I figure everybody has an off day.
- I'll have more grades than I know what to do with.
- I give quizzes and tests and I work on a point-total system
- I would break the quizzes, tests, and homework into a percentage grade.

Borderline Grades

Every teacher faces decisions concerning grades that are borderline, just between two letter grades, or very close to a higher grade. Teachers in this sample indicated that in these situations they want to be able to give students the benefit of the doubt (pulling for students), and typically use non-achievement factors for making their decision, such as effort and participation, or use extra work or extra credit. This reflects the teachers' desire for students to be as successful as possible and to obtain the highest grade possible. It is usually a subjective judgment by the teacher. Here are some illustrations of what teachers do with borderline grades:

- I will suggest that maybe they do something extra, which could be a project … I tutor with them … I'll give them make-up work because usually they don't even ask for make-up work they missed. It depends on the situation, but I do what I can to try to help them over the hump.
- Borderline comes down to effort.
- Borderline, effort is the key … can make up zeros or use extra credit.
- If they come in and say they got a 60 the first time and they come in and get a 85, then I'll up that to a 75 or something.
- An F is a 63, those kids get 60. I will pass them especially if they've really showed me the effort … if I know they're really trying and I mean, genuinely, then I will pass them.
- If I'm within a point or a point and a half of the next letter grade, I look at the child and do I feel the child has made an effort?
- Then I generally think of their effort, whether I feel they've really tried and whether they've turned in all their work. If they didn't turn it all in and it's borderline, I don't give it to them … if they tried to make an effort to improve3, I won't give them an F; if

they didn't do their work and they've been absent quite a bit, then they're gonna get what they deserve.

- When it's borderline, how hard has the child worked in the year? And I will be honest with them, it it's a 63.5, I'm going to bring it up to 64.
- Borderline, most of the time I round up … I'll give extra points to someone who really works hard.
- Reserve A for performance, B for effort is possible.

There were a few teachers who clearly did not want to use subjective criteria for borderline situations:

- Frankly I tell them that when they get report cards they will be abased oh objectives and how well people meet them. How can I grade on effort?
- The calculator decides [borderline cases] … to me, I round up half a point … I try to set up the system where I don't have to make evaluative judgments.

Student Effort

One of the most varied practices in grading students is concerned with how the teacher recognizes and handles student effort. From one standpoint, most teachers use effort to some extent in deciding borderline cases, giving a student who tries hard the higher grade. Many teachers view effort as enabling achievement or as part of achievement, so that it becomes and important contributor to determining grades. Some teachers do not use effort at all, relying instead solely on the quality of student performance. Many teachers think of homework as a proxy for effort. The following quotes show how different teachers have different ideas about how to handle effort.

- At this level you have to take into consideration effort but it can't be to the exclusion of performance because it's a fine line.
- I have one child I think is getting a D and she had worked like a dog and so we really just bumped her up to a C.
- As far as an effort grade, I don't really believe in effort grades but, well, homework is a good example, I give an effort grade for having homework everyday.
- Most projects there is usually a window where I'm grading effort. I can tell that some have been working really hard and I'm going to give them the benefit of the doubt

…there is one girl that tries really hard and all she can get is a high F, and I give her a D every time. I will not fail that girl.

- I want to see that there were sincere efforts. When I look and see that a child's missed eight out of ten homework assignments … he decided just to sit there and not do them … that's what I measure as not sincere efforts.
- I put effort in their class participation grade. Some students sit there and don't say a word. I factor in not only their actual class participation, but also their effort, what I perceive is effort.
- So to me, conduct and behavior and attendance is very, very important in assessing that final grade.
- It [effort] only comes into play in that test and essay realm and then in the end result … if Johnny probably deserved to have an 83, I would maybe for that effort give him above for the grade … and on the other side of that coin, I would maybe not bump her down.

Extra Credit

Teachers were asked how they incorporated extra credit in grading students, if at all. Most teachers do use extra credit, primarily as a way of boosting the grades of students who may be borderline or receiving a low grade (pulling for students). There are many different ways extra credit is used. Some teachers make is relatively easy for students and have an informal set of guidelines, while other teachers believe students must clearly earn the extra credit with additional effort. Another variable with extra credit is whether students know about it and can plan for it, or whether the teacher simply awards extra credit as a surprise. Both approaches are used. Many teachers offer ways to earn extra points as extra credit. Some comments of teachers about extra credit are the following:

- I tell them they can have extra credit when they have done what they have supposed to do for credit. Make them learn and see that they have to put forth some effort. I think too many kids get by today with not earning what they get , and that's an important lesson.
- If somebody does extra credit and it doesn't indicate better performance, then no I'm not going to give them anything. There's no like free points. I always retest everybody who gets Ds and Fs, and I'll throw out the old one. I will always give people a chance to improve, why not?
- The things that motivate my kids is they'll put so much more effort into the extra credit than they will the regular work. They love to see 75 + 10.

- If it was a particularly hard homework question, the ones who got that when I go around … I give them extra credit. They never knew that part ahead of time.
- I rather them do the assignment themselves rather than give extra credit. But what I do is offer bonus points, which, I guess is almost the same. For example, just things like taking home papers to get signed if they bring them back.
- It's worth a few extra points if they're willing to show things to their parents to keep them abreast of what's going on in the classroom.
- I don't give extra credit. I tell students that you earn your grade … you don't come in at the last minute and ask for a bail out … [but] we do have extra assignments that are optional that you can do to earn extra points.
- They get two make-up assignments; that's the extra credit.
- Sometimes I'll give them an extra credit problem or a project or something like that.
- I have one class when they have to bring their report cards back signed they'll get 3 or 4 points.

Handling Zeros

A vexing problem in grading students is how to handle zeros. Our teachers reported a variety of ways that zeros are used. Teachers generally understand the devastating effect a zero can have on grades, and most teachers try to accommodate students by providing opportunities to remove zeros (pulling for students). Some teachers use zeros for motivation. Generally a zero is intended for no work at all, not for receiving an F. Like other assessment and grading practices, zeros are handled in ways that make most sense for individual students, despite the presence of a single policy.

- A zero means you didn't do the project at all … an F means you did the work and you deserve some credit. For the most part I try not to let the kids get a D or F, I have what you call do-overs.
- If they [students] just got one zero, I mean I'm lenient enough. They are not going to figure out the percentages anyway, so I can fix it then.
- I put the zero in at the end of the nine weeks if they just haven't turned anything in … I try to make sure they have an opportunity to make it up. I know a zero will kill their grade and they don't understand that.
- I have a lot of grades, so one zero does not make a great deal of difference … it's all done in percentages so at the beginning it has a heavy effect … I do not give them a chance to make it up.

- It [zero] counts as a regular grade. One of the things I discuss when we are covering means is that every zero counts, don't miss assignments and think you're getting over, you're not.
- Oh, I record them [zeros] to start with, but I don't know. Maybe this is the dedicated teacher's syndrome or whatever. I'll chase the kid around for a long time so that I can get a few points. I have a child now who is absolutely and A+ student. She hasn't turned in her last writing assignment … so it's dropped her A+ to a C- … I've hounded her every single day.
- I don't give a zero … it's murder for a child to make up. There are people that give 0's and it just turns the kids right off.
- I cannot change a grade … the zeros stay there … the zero stays if they don't make it up … there's a lot of stuff I want to broadcast, but I just can't turn them down when a kid comes to me.
- If it is a graded assignment then yes, I consider it a zero, but I offer them an opportunity to go back and o them. It's the learning that I'm most interested in, not the penalizing.
- It's very straightforward. They are just average in and if there are mitigating circumstances then I would take that into consideration.

Conclusions

The results of these analyses indicates that teachers have a lot to say about their idiosyncratic assessment and grading practices. It appears that teachers are constantly striving to reach a reasonable balance between their beliefs about education and learning on the one hand, with the pressures exerted by external factors. This constant state of tension may help explain why teachers view assessment and grading as a fluid set of principles that change to some extent each year. Together, these influences converge on the actual process of making assessment and grading decision, which result in turn in the development and implementation of assessment and grading practices. Because of the interplay between the teachers' beliefs, external factors, and student characteristics, a great amount of variety in classroom assessment and grading is evident.

Important teacher beliefs that influence decision-making include a larger philosophy of teaching and learning, wanting students to succeed, accommodating individual differences among students, engaging and motivating students to learn, and promoting student understanding

and mastery. These beliefs converge on getting students, in whatever ways are necessary, to be involved in learning, giving effort, and ultimately demonstrating successful performance.

Important external forces include the SOL and SOL tests, district grading policies, and parents. Clearly, the most important external factor are the SOL and SOL tests. These externally mandated high stakes tests have put pressure on teachers to modify their assessment practices to accommodate the SOL and the format of the SOL tests.

One impression is the strong sense of ownership teachers have for their assessment and grading practices. It is almost as if there is a sense of pride and ownership that the practices are unique and that they have a good rationale for them. It also seems that assessment and grading is largely a private business, not readily talked about very much with other teachers. Clearly, assessment and grading practices fit within a larger philosophy of student learning, and clearly teachers are very interested in and committed to enhancing the learning of each student. They want students to learn. So it follows that they want assessment and grading practices to enhance student learning, not simply document student performance.

Assessment practices that emerged from the interviews stressed the wide variety of assessments used for different purposes, and the need for variety to accommodate student learning styles. Formative assessments are used constantly during instruction to inform teaching decisions. Pre-assessments are sometimes used prior to instruction to gauge current student knowledge. Revisions are made continuously by teachers, and teachers, in the main, construct the assessments they use with their students.

Grading practices are very idiosyncratic. Teachers adopt their own grading policy, with little regard for standardization with other teachers. Most teachers use effort as a determining factor in borderline grades, and in general believe that student effort is a good proxy for student

achievement. Extra credit is used to help students obtain a higher grade. There is great variety in how zeros are handled.

An important finding from these data is that classroom assessment and grading are integrated with instruction. Most teachers see assessment and grading as extensions of instruction that have important consequences for student engagement and motivation. Thus, teachers' decision-making is heavily influenced by thinking about how the assessments will enhance student learning. Teachers believe that learning is best assessed with multiple assessments, using different formats. They also believe that informal or formative, and constructed-response assessments provide the best information to judge student understanding.

Our goal in this study was to "get inside the head" of teachers to find out what influences their decision-making concerning assessment and grading practices. We have learned that decisions are made on the basis of how the assessment or grading procedure will affect student learning and motivation, and, at the same time, respond to external pressures. In this balancing act each teacher has his or her own solution, one that is constantly changing with each new group of students.

Implications

The results of this study suggest several implications. First, given that teachers clearly "pull" for student success and use many different practices that help student succeed, it may be helpful to ask if teachers are "coddling" students, making it so easy to obtain passing and even high grades that students are getting a false sense of their own level of understanding and performance. In other words, is the desire of the teacher to see student "success" so strong that it

promotes assessment and grading practices that students can obtain good grades without really knowing the content or being able to demonstrate the skill?

Second, what are the results of emphasizing effort as much as teachers do in grading students? Research on student motivation and attributions for success (reasons students give for their success) suggest that while an emphasis on effort is positive for motivation because effort is a controllable, internal factor, it may be counterproductive for some low performing students because they may develop a belief that they can be rewarded for effort and not mastery of the content or skills. This may also give students a false sense of their competence. Furthermore, too great an emphasis on effort may reduce attributions to ability, which are more stable. On the other hand, this emphasis on effort at least teaches students the importance of engagement and involvement and the need for this involvement to be successful.

A third implication is concerned with the skills teachers have to construct and revise classroom assessments. It is clear from the literature, and the results from Phase 1 of this research effort, that teachers may not have the knowledge and skills that are needed to effectively construct and revise assessments. With the popularity of new types of assessments, such as performance and portfolio assessments, teacher skills in assessment may be thinned even further. It may be helpful to systematically evaluate teachers' assessment skills and provide professional development where needed.

A fourth implication of these findings is the potential effect of external pressures on teacher professionalism. The influence of the SOL and SOL tests is undeniable, and seems to be directed at something that is very important to teachers' sense of what it means to be an effective teacher. Teachers desire autonomy and need to adapt instruction and assessment to their personal styles and to the needs of individual students. Teachers do not appreciate standardization of practices

that minimize these dimensions of being a teacher, and the SOL and SOL tests have had such a standardization effect. The question is whether this, in fact, is affecting teachers' sense of professionalism, and if so, what impact this has on teacher morale and motivation. In addition, it may be that in Virginia, at this time, external pressures are particularly influential given the current situation with the SOL testing.

A fifth implication concerns teacher training and teacher induction. What do these data suggest with respect to how teachers are trained? How important is it for teachers to have a fully developed philosophy of teaching and learning so that assessment and grading practices can be based on this philosophy? What is being done in teacher training to help teachers become competent in the variety of assessment methods that are typically used, as well as how to integrate external pressures with personal beliefs and district grading policies? In the induction of beginning teachers it may be valuable to examine their assessment and grading practices to see if they are consistent with philosophy of teaching and learning and other beliefs. For example, if a strong value in teaching in maximizing the learning of each student, what adjustments in assessments are made to accommodate individual differences among students?

It is clear that teachers spend a great deal of time with assessment and grading, and that they see these tasks as integral to the teaching/learning process. This research helps to show how teachers make assessment and grading decisions, pointing to tension teachers feel when internal beliefs and values conflict with external pressures and demands. This understanding will hopefully suggest positive actions that can improve assessment and grading practices.

References

Airasian, P. W. (1997). Classroom assessment (3rd Edition). NY: McGraw-Hill.

Airasian, P. W. (1984). Classroom assessment and educational improvement. Paper presented at the conference Classroom Assessment: A Key to Educational Excellence, Northwest Regional Educational Laboratory, Portland, Oregon.

American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, NEA). (1990) Standards for teacher competence in educational measurement. Washington, DC:. Author.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology, 84, 261-271.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. Applied Measurement in Education, 10, 161-180.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. Applied Measurement in Education, 7, 279-301.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. Journal of Educational Measurement, 30, 123-142.

Brookhart, S. M. (1991). Grading practices and validity. Educational Measurement: Issues and Practice, 10, 35-36.

Carter, K. (1984). Do teachers understand the principles for writing tests? Journal of Teacher Education, 35, 57-60.

Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye, Editor, Handbook of classroom assessment, NY: Academic Press.

Cizek, G. J., Fitzgerald, Shawn M., & Rachor, Robert E. (1995). Teachers' Assessment Practices: Preparation, Isolation and the Kitchen Sink. Educational Assessment, 3(2), 159-179.

Cross, L. H., & Frary, R. B. (1996). Hodgepodge grading: Endorsed by students and teachers alike. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway, ed., Testing in the schools. New directions for testing and measurement. San Francisco: Jossey-Bass.

Frary, R.B., Cross, L.H. & Weber, L.J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice 12(3), 23-30.

Friedman, S. J. & Manley, M. (1991). Grading practices in the secondary school: Perceptions of the stakeholders. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago.

Haertel, E., Ferrara, S., Korpi, M., & Prescott, B. (1984). Testing in secondary schools: Student perspectives. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

McMillan, J. H. (1997). Classroom assessment: Principles and practice for effective instruction. Boston: Allyn & Bacon.

Messick, S. (1989). Validity. In R. L. Linn (ed.), Educational measurement (3rd Edition). Washington, DC: American Council on Education and National Council on Measurement in Education.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62, 229-258.

Phye, G. D. (1997). Classroom assessment: A multidimensional perspective. In G. D. Phye, Editor, Handbook of classroom assessment, NY: Academic Press.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye, ed., Handbook of classroom assessment, NY: Academic Press.

Plake, B. S., & Impara, J. C. (1993). Assessment competencies of teachers: A national survey. Educational Measurement: Issues and Practice, 12, 10-25.

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Shunk & J. L. Meece (Eds.), Student perceptions in the classroom (pp. 149-183). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pintrich, P. R., & Schunk, D. H. (1996). Motivation in education: Theory, research, and applications, Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J. (1995). Classroom assessment: What teachers need to know, Boston: Allyn & Bacon.

Schafer, W. D. (1993). Assessment in teacher education. Theory into Practice, 32, 118-126.

Schunk, D. H. (1994). Self-regulation of self-efficacy and attributions in academic settings. In D. H. Schunk & B. J. Zimmerman (Eds.), Self-regulation of learning and performance: Issues and educational applications (pp. 75-99). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shulman, L. S. (1980). Test design: A view from practice. In E. L. Baker and E. S. Quellmalz, (Eds.), Educational testing and evaluation. Los Angeles: Sage.

Stiggins, R. J. (1997). Student-centered classroom assessment. Upper Saddle River, NJ: Prentice-Hall.

Stiggins, R. J. & Conklin, N. F. (1992). In teacher's hands: Investigating the practices of classroom assessment, Albany, NY: State University of New York.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school: Building a research agenda. Educational Measurement: Issues and Practice, 8, 5-14.

Stipek, D. (1998). Motivation to learn: From theory to practice. 3rd ed., Boston: Allyn & Bacon.

Tittle, C. K. (1994). Toward an educational psychology of assessment for teaching and learning: Theories, contexts, and validation arguments. Educational Psychologist, 29, 149-162.

Truog, A. L. & Friedman, S. J. (1996). Evaluating high school teachers' written grading policies from a measurement perspective. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. Psychological Review, 92(4), 548-73.

Appendix A

Interview Protocol

Topic Guide

2/1/98

Classroom Assessment and Grading Teacher Interview

Directions: The purpose of this Interview Topic Guide is to provide a protocol for asking questions to elicit teacher responses concerning their classroom assessment and grading practices. The general purpose of the interview is to obtain a thorough understanding of **why** teachers use certain assessment and grading practices; **their reasons and decision-making** concerning these practices. It is important for the teachers to be informed that their responses are completely confidential; and they should be encouraged to be as honest as possible. Use the following questions as a guide and make notes of responses in the space provided. After asking the first two questions, feel free to use whatever order seems best in asking the questions, and use prompts as needed.

Begin the interview with "small talk" and other conversation to put the teacher at ease and create a comfortable environment. Define 'classroom assessment' as policies, techniques, and procedures used to measure, interpret, and use information to make decisions about what students know and can do on units, chapters, and other major learning goals (e.g., tests, quizzes, homework, papers, reports, etc.).

1. To begin our interview, I'd like you to select a single class you are currently teaching that would be an example of the **most typical class** you usually teach. Then I would like you to answer <u>all</u> of the following questions based on this class. Please describe the class for me, with respect to:

   Grade level: _____
   Subject: _____
   Ability level: _____
   Class size: _____

   Also, how many years have you taught at this level? _____

2. Briefly, please describe the kinds of assessments that you use for this class.

3. **When** do you decide which assessments to use?

   Probe: Before the class begins or during the class?

4. **Why** do you use these particular kinds of assessments?

   Probes: Do you ever change assessments? If so, why?
   Do you use them because that's what other teachers use?
   Do you use them because it is what is suggested in instructional materials?
   Do you use them because they motivate students?
   Do you use them because of tradition?
   Do you use them to provide an "objective" record of student performance?
   Do you depend on the school's policy, or is it your decision?

5. Which of your assessments do you find to be most valuable for informing you about how much students know and can do? Why?

   Probe: Do you use tests already prepared by the department or publisher?

6. How do you think classroom assessments, like papers, tests, and other assignments affect student motivation? What kinds of assessments seem to motivate students more than other kinds? Please explain why.

   Probe: Are your expectations communicated the difficulty of the tests?
   How is student effort assessed?

7. Briefly, what is your grading policy, and **how did you come to decide** what it would be?

   Probes: How do you incorporate student effort?
   How do you handle borderline grades?
   How do you use extra credit?
   How do you handle zeros?

8. What is your typical distribution (or spread) of grades given?

   Probes: How did you come to believe that that distribution was appropriate?
   Do you ever talk to other teachers about grade distributions?

9. Do you pre-assess students, either formally or informally, to determine their strengths and weaknesses?

   Probes: If so, how and how often?

If not, how do you gage student knowledge and skills prior to instruction?
Is pre-assessment used to better plan instruction to meet student needs?

10. Do your lesson plans determine the assessments you use, or do your assessments dictate your lesson plans?

Probe: Do you use test results to re-instruct students on weaknesses demonstrated through their performances?

11. Do you modify assessments on the basis of student characteristics?

Probes: Do you offer different tests to students at different ability levels?
Different types of assessments?
Make modifications?

12. When do you typically write and revise tests that students take?

Probes: How often are major tests revised?
To what extent are they revised?
Why are the tests revised?

13. What kind of assessments, either formal or informal, do you use day to day to inform you about how much students know and how much progress they are making?

14. How do you give feedback to students when returning an assignment or test?

Probes: Is it done individually or as a group?
Why do you use this kind of feedback and not use other kinds?
How do you handle feedback when a student has failed?

15. How do you think the new SOL tests will influence your classroom assessments?

16. Use the following scale to answer a few questions about factors that contribute to semester grades you will give to the class you have described. After using the scale provided, then estimate the percentage that factor contributed to the final grade.

1	2	3	4	5	6
Not at all	Very Little	Some	Quite A Bit	Extensively	Completely

2a. **student effort – how much the student tried to learn**
What percentage of the final semester grade is based on student effort?

2b. **assessments that measure student recall knowledge (content)**
What percentage of the final semester grade is based on assessments that measure recall knowledge?

2c. **performances on quizzes**
What percentage of the final semester grade is based on performances on quizzes?

2e. **objective assessments** (e.g., matching, multiple choice, short answer)
What percentage of the final semester grade is based on objective assessments?

(tear off and give to teacher)

1	2	3	4	5	6
Not at all	Very Little	Some	Quite A Bit	Extensively	Completely